A-67991-2

DETECTION OF NUCLEIC ACID AMPLIFICATION REACTIONS USING BEAD ARRAYS

SUB B1

The present invention is a continuation-in-part of U.S.S.N. 60/161,148, filed October 22, 1999, which is a continuation-in-part of U.S.S.N. 60/135,051, filed May 20, 1999, and a continuation-in -part of U.S.S.N. 60/160,027, filed October 22, 1999, which is a continuation-in-part of U.S.S.N. 60/130,089, filed April 20, 1999, all of which are pending.

FIELD OF THE INVENTION

The invention relates to compositions and methods useful in the detection and quantification of a nucleic acid target using a variety of amplification techniques, including both signal amplification and target amplification. Detection proceeds through the use of a label that is associated with the amplified signal or target, either directly or indirectly, to allow optical detection of the light absorbing label using a microsphere array sensor.

BACKGROUND OF THE INVENTION

The detection of specific nucleic acids is an important tool for diagnostic medicine and molecular biology research. Gene probe assays currently play roles in identifying infectious organisms such as bacteria and viruses, in probing the expression of normal genes and identifying mutant genes such as oncogenes, in typing tissue for compatibility preceding tissue transplantation, in matching tissue or blood samples for forensic medicine, and for exploring homology among genes from different species.

Ideally, a gene probe assay should be sensitive, specific and easily automatable (for a review, see Nickerson, Current Opinion in Biotechnology 4:48-51 (1993)). The requirement for sensitivity (i.e. low detection limits) has been greatly alleviated by the development of the polymerase chain reaction (PCR) and other amplification technologies which allow researchers to amplify exponentially a specific nucleic acid sequence before analysis as outlined below (for a review, see Abramson et al., Current Opinion in Biotechnology, 4:41-47 (1993)).

Sensitivity, i.e. detection limits, remain a significant obstacle in nucleic acid detection systems, and a variety of techniques have been developed to address this issue. Briefly, these techniques can be classified as either target amplification or signal amplification. Target amplification involves the amplification (i.e. replication) of the target sequence to be detected, resulting in a significant increase in the number of target molecules. Target amplification strategies include the polymerase chain

reaction (PCR), strand displacement amplification (SDA), and nucleic acid sequence based amplification (NASBA).

Alternatively, rather than amplify the target, alternate techniques use the target as a template to replicate a signalling probe, allowing a small number of target molecules to result in a large number of signalling probes, that then can be detected. Signal amplification strategies include the ligase chain reaction (LCR), cycling probe technology (CPT), invasive cleavage techniques such as Invader™ technology, Q-Beta replicase (QβR) technology, and the use of "amplification probes" such as "branched DNA" that result in multiple label probes binding to a single target sequence.

The polymerase chain reaction (PCR) is widely used and described, and involves the use of primer extension combined with thermal cycling to amplify a target sequence; see U.S. Patent Nos. 4,683,195 and 4,683,202, and PCR Essential Data, J. W. Wiley & sons, Ed. C.R. Newton, 1995, all of which are incorporated by reference. In addition, there are a number of variations of PCR which also find use in the invention, including "quantitative competitive PCR" or "QC-PCR", "arbitrarily primed PCR" or "AP-PCR" , "immuno-PCR", "Alu-PCR", "PCR single strand conformational polymorphism" or "PCR-SSCP", allelic PCR (see Newton et al. Nucl. Acid Res. 17:2503 91989); "reverse transcriptase PCR" or "RT-PCR", "biotin capture PCR", "vectorette PCR". "panhandle PCR", and "PCR select cDNA subtraction", among others.

Strand displacement amplification (SDA) is generally described in Walker et al., in Molecular Methods for Virus Detection, Academic Press, Inc., 1995, and U.S. Patent Nos. 5,455,166 and 5,130,238, all of which are hereby incorporated by reference.

Nucleic acid sequence based amplification (NASBA) is generally described in U.S. Patent No. 5,409,818 and "Profiting from Gene-based Diagnostics", CTB International Publishing Inc., N.J., 1996, both of which are incorporated by reference.

Cycling probe technology (CPT) is a nucleic acid detection system based on signal or probe amplification rather than target amplification, such as is done in polymerase chain reactions (PCR). Cycling probe technology relies on a molar excess of labeled probe which contains a scissile linkage of RNA. Upon hybridization of the probe to the target, the resulting hybrid contains a portion of RNA:DNA. This area of RNA:DNA duplex is recognized by RNAseH and the RNA is excised, resulting in cleavage of the probe. The probe now consists of two smaller sequences which may be released, thus leaving the target intact for repeated rounds of the reaction. The unreacted probe is removed and the label is then detected. CPT is generally described in U.S. Patent Nos. 5,011,769, 5,403,711, 5,660,988, and 4,876,187, and PCT published applications WO 95/05480, WO 95/1416, and WO 95/00667, all of which are specifically incorporated herein by reference.

The oligonucleotide ligation assay (OLA; sometimes referred to as the ligation chain reaction (LCR)) involve the ligation of at least two smaller probes into a single long probe, using the target sequence as the template for the ligase. See generally U.S. Patent Nos. 5,185,243, 5,679,524 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; and WO 89/09835, all of which are incorporated by reference.

Invader™ technology is based on structure-specific polymerases that cleave nucleic acids in a site-specific manner. Two probes are used: an "invader" probe and a "signalling" probe, that adjacently hybridize to a target sequence with a non-complementary overlap. The enzyme cleaves at the overlap due to its recognition of the "tail", and releases the "tail" with a label. This can then be detected. The Invader™ technology is described in U.S. Patent Nos. 5,846,717; 5,614,402; 5,719,028; 5,541,311; and 5,843,669, all of which are hereby incorporated by reference.

"Rolling circle amplification" is based on extension of a circular probe that has hybridized to a target sequence. A polymerase is added that extends the probe sequence. As the circular probe has no terminus, the polymerase repeatedly extends the circular probe resulting in concatamers of the circular probe. As such, the probe is amplified. Rolling-circle amplification is generally described in Baner *et al.* (1998) *Nuc. Acids Res.* 26:5073-5078; Barany, F. (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193; and Lizardi *et al.* (1998) *Nat. Genet.* 19:225-232, all of which are incorporated by reference in their entirety.

"Branched DNA" signal amplification relies on the synthesis of branched nucleic acids, containing a multiplicity of nucleic acid "arms" that function to increase the amount of label that can be put onto one probe. This technology is generally described in U.S. Patent Nos. 5,681,702, 5,597,909, 5,545,730, 5,594,117, 5,591,584, 5,571,670, 5,580,731, 5,571,670, 5,591,584, 5,624,802, 5,635,352, 5,594,118, 5,359,100, 5,124,246 and 5,681,697, all of which are hereby incorporated by reference.

Similarily, dendrimers of nucleic acids serve to vastly increase the amount of label that can be added to a single molecule, using a similar idea but different compositions. This technology is as described in U.S. Patent No. 5,175,270 and Nilsen et al., J. Theor. Biol. 187:273 (1997), both of which are incorporated herein by reference.

In each of these methods, analysis of the amplified target or amplified signal remains problematic. Accordingly, it is an object of the invention to provide compositions and methods for the detection and quantification of the products, either directly or indirectly, of nucleic acid amplification reactions.

SUMMARY OF THE INVENTION

In accordance with the objects outlined above, the present invention provides a method for detecting a

first target nucleic acid sequence. In one aspect the method comprises hybridizing at least a first primer nucleic acid to the first target sequence to form a first hybridization complex, contacting the first hybridization complex with a first enzyme to form a modified first primer nucleic acid, disassociating the first hybridization complex, contacting the modified first primer nucleic acid with an array comprising a substrate with a surface comprising discrete sites and a population of microspheres comprising at least a first subpopulation comprising a first capture probe such that the first capture probe and the modified primer form an assay complex, wherein the microspheres are distributed on the surface, and detecting the presence of the modified primer nucleic acid.

In addition the method further comprises hybridizing at least a second primer nucleic acid to a second target sequence that is substantially complementary to the first target sequence to form a second hybridization complex, contacting the second hybridization complex with the first enzyme to form modified second primer nucleic acid, disassociating the second hybridization complex and forming a second assay complex comprising the modified second primer nucleic acid and a second capture probe on a second subpopulation.

In an additional aspect of the invention the primer forms a circular probe following hybridization with the target nucleic acid to form a first hybridization complex and contacting the first hybridization complex with a first enzyme comprising a ligase such that the oligonucleotide ligation assay (OLA) occurs. This is followed by adding the second enzyme, a polymerase, such that the circular probe is amplified in a rolling circle amplification (RCA) assay.

In an additional aspect of the invention, the first enzyme comprises a DNA polymerase and the modification is an extension of the primer such that the polymerase chain reaction (PCR) occurs. In an additional aspect of the invention the first enzyme comprises a ligase and the modification comprises a ligation of the first primer which hybridizes to a first domain of the first target sequence, to a third primer which hybridizes to a second adjacent domain of the first target sequence such that the ligase chain reaction (LCR) occurs.

In an additional aspect of the invention, the first primer comprises a first probe sequence, a first scissile linkage and a second probe sequence, wherein the first enzyme will cleave the scissile linkage resulting in the separation of the first and second probe sequences and the disassociation of the first hybridization complex, leaving the first target sequence intact such that the cycling probe technology (CPT) reaction occurs.

In addition, wherein the first enzyme is a polymerase that extends the first primer and the modified first primer comprises a first newly synthesized strand, the method further comprises the addition of a second enzyme comprising a nicking enzyme that nicks the extended first primer leaving the first target sequence intact, and extending from the nick using the polymerase, and thereby displacing the

first newly synthesized strand and generating a second newly synthesized strand such that strand displacement amplification (SBA) occurs.

In addition, wherein the first target sequence is an RNA target sequence, the first primer nucleic acid is a DNA primer comprising an RNA polymerase promoter, the first enzyme is a reverse-transcriptase that extends the first primer to form a first newly synthesized DNA strand, the method further comprises the addition of a second enzyme comprising an RNA degrading enzyme that degrades the first target sequence, the addition of a third primer that hybridizes to the first newly synthesized DNA strand, the addition of a third enzyme comprising a DNA polymerase that extends the third primer to form a second newly synthesized DNA strand, to form a newly synthesized DNA hybrid, the addition of a fourth enzyme comprising an RNA polymerase that recognizes the RNA polymerase promoter and generates at least one newly synthesized RNA strand from the DNA hybrid, such that nucleic acid sequence-based amplification (NASBA) occurs.

In addition, wherein the first primer is an invader primer, the method further comprises hybridizing a signalling primer to the target sequence, the enzyme comprises a structure-specific cleaving enzyme and the modification comprises a cleavage of said signalling primer, such that the invasive cleavage reaction occurs.

An additional aspect of the invention is a method for detecting a target nucleic acid sequence comprising hybridizing a first primer to a first target sequence to form a first hybridization complex, contacting the first hybridization complex with a first enzyme to extend the first primer to form a first newly synthesized strand and form a nucleic acid hybrid that comprises an RNA polymerase promoter, contacting the hybrid with an RNA polymerase that recognizes the RNA polymerase promoter and generates at least one newly synthesized RNA strand, contacting the newly synthesized RNA strand with an array comprising a substrate with a surface comprising discrete sites and a population of microspheres comprising at least a first subpopulation comprising a first capture probe; such that the first capture probe and the modified primer form an assay complex; wherein the microspheres are distributed on the surface and detecting the presence of the newly synthesized RNA strand.

In addition, when the target nucleic acid sequence is an RNA sequence, and prior to hybridizing a first primer to a first target sequence to form a first hybridization complex, method comprises hybridizing a second primer comprising an RNA polymerase promoter sequence to the RNA sequence to form a second hybridization complex, contacting the second hybridization complex with a second enzyme to extend the second primer to form a second newly synthesized strand and form a nucleic acid hybrid; and degrading the RNA sequence to leave the second newly synthesized strand as the first target sequence. In one aspect of the invention the degrading is done by the addition of an RNA degrading enzyme. In an additional aspect of the invention the degrading is done by RNA degrading activity of reverse transcriptase.

In addition, when the target nucleic acid sequence is a DNA sequence, and prior to hybridizing a first primer to a first target sequence to form a first hybridization complex, the method comprises hybridizing a second primer comprising an RNA polymerase promoter sequence to the DNA sequence to form a second hybridization complex, contacting the second hybridization complex with a second enzyme to extend the second primer to form a second newly synthesized strand and form a nucleic acid hybrid, and denaturing the nucleic acid hybrid such that the second newly synthesized strand is the first target sequence.

An additional aspect fo the invention is a kit for the detection of a first target nucleic acid sequence. The kit comprises at least a first nucleic acid primer substantially complementary to at least a first domain of the target sequence, at least a first enzyme that will modify the first nucleic acid primer, and an array comprising a substrate with a surface comprising discrete sites, and a population of microspheres comprising at least a first and a second subpopulation, wherein each subpopulation comprises a bioactive agent, wherein the microspheres are distributed on the surface.

In an additional aspect of the invention, is a kit for the detection of a PCR reaction wherein the first enzyme is a thermostable DNA polymerase.

In an additional aspect of the invention, is a kit for the detection of a LCR reaction wherein the first enzyme is a ligase and the kit comprises a first nucleic acid primer substantially complementary to a first domain of the first target sequence and a third nucleic acid primer substantially complementary to a second adjacent domain of the first target sequence.

In an additional aspect of the invention, is a kit for the detection of a strand displacement amplification (SDA) reaction wherein the first enzyme is a polymerase and the kit further comprises a nicking enzyme.

In an additional aspect of the invention, is a kit for the detection of a NASBA reaction wherein the first enzyme is a reverse transcriptase, and the kit comprises a second enzyme comprising an RNA degrading enzyme, a third primer, a third enzyme comprising a DNA polymerase and a fourth enzyme comprising an RNA polymerase.

In an additional aspect of the invention, is a kit for the detection of an invasive cleavage reaction wherein the first enzyme is a structure-specific cleaving enzyme, and the kit comprises a signaling primer.

DETAILED DESCRIPTION OF THE FIGURES

Figures 1A, 1B and 1C depict three different embodiments for attaching a target sequence to an array.

The solid support **5** has microsphere **10** with capture probe **20** linked via a linker **15**. Figure 1A depicts direct attachment; the capture probe **20** hybridizes to a first portion of the target sequence **25**. Figure 1B depicts the use of a capture extender probe **30** that has a first portion that hybridizes to the capture probe **20** and a second portion that hybridizes to a first domain of the target sequence **25**. Figure 1C shows the use of an adapter sequence **35**, that has been added to the target sequence, for example during an amplification reaction as outlined herein.

Figures 2A and 2B depict two preferred embodiments of SBE amplification. Figure 2A shows extension primer **40** hybridized to the target sequence **25**. Upon addition of the extension enzyme and labelled nucleotides, the extension primer is modified to form a labelled primer **41**. The reaction can be repeated and then the labelled primer is added to the array as above. Figure 2B depicts the same reaction but using adapter sequences.

Figures 3A and 3B depict two preferred embodiments of OLA amplification. Figure 3A depicts a first ligation probe **45** and a second ligation probe **50** with a label **55**. Upon addition of the ligase, the probes are ligated. The reaction can be repeated and then the ligated primer is added to the array as above. Figure 3B depicts the same reaction but using adapter sequences.

Figure 4 depicts a preferred embodiment of the invasive cleavage reaction. In this embodiment, the signaling probe **65** comprises two portions, a detection sequence **67** and a signaling portion **66**. The signaling portion can serve as an adapter sequence. In addition, the signaling portion generally comprises the label **55**, although as will be appreciated by those in the art, the label may be on the detection sequence as well. In addition, for optional removal of the uncleaved probes, a capture tag **60** may also be used. Upon addition of the enzyme, the structure is cleaved, releasing the signaling portion **66**. The reaction can be repeated and then the signaling portion is added to the array as above.

Figures 5A and 5B depict two preferred embodiments of CPT amplification. A CPT primer **70** comprising a label **55**, a first probe sequence **71** and a second probe sequence **73**, separated by a scissile linkage **72**, and optionally comprising a capture tag **60**, is hybridized to the target sequence **25**. Upon addition of the enzyme, the scissile linkage is cleaved. The reaction can be repeated and then the probe sequence comprising the label is added to the array as above. Figure 5B depicts the same reaction but using adapter sequences.

Figure 6 depicts OLA/RCA amplification using a single "padlock probe" **57**. The padlock probe is hybridized with a target sequence **25**. When the probe **57** is complementary to the target sequence **26**, ligation of the probe termini occurs forming a circular probe **28**. When the probe **57** is not complementary to the target sequence **27**, ligation does not occur. Addition of polymerase and nucleotides to the circular probe results amplification of the probe **58**. Cleavage of the amplified probe

SUB B3 Cont. 58 yields fragments **59** that hybridize with an identifier probe **21** immobilized on a microsphere **10**.

Figure 7 depicts an alternative method of OLA/RCA. An immobilized first OLA primer **45** is hybridized with a target sequence **25** and a second OLA primer **50**. Following the addition of ligase, the first and second OLA primers are ligated to form a ligated oligonucleotide **56**. Following denaturation to remove the target nucleic acid, the immobilized ligated oligonucleotide is distributed on an array. An RCA probe **57** and polymerase are added to the array resulting in amplification of the circular RCA probe **58.**

SUB B4

DETAILED DESCRIPTION OF THE INVENTION

This invention is directed to the detection (and optionally quantification) of products of nucleic acid amplification reactions, using bead arrays for detection of the amplification products. Suitable amplification methods include both target amplification and signal amplification and include, but are not limited to, polymerase chain reaction (PCR), ligation chain reaction (sometimes referred to as oligonucleotide ligase amplification OLA), cycling probe technology (CPT), strand displacement assay (SDA), transcription mediated amplification (TMA), nucleic acid sequence based amplification (NASBA), rolling circle amplification (RCA), and invasive cleavage technology. All of these methods require a primer nucleic acid (including nucleic acid analogs) that is hybridized to a target sequence to form a hybridization complex, and an enzyme is added that in some way modifies the primer to form a modified primer. For example, PCR generally requires two primers, dNTPs and a DNA polymerase; LCR requires two primers that adjacently hybridize to the target sequence and a ligase; CPT requires one cleavable primer and a cleaving enzyme; invasive cleavage requires two primers and a cleavage enzyme; etc. Thus, in general, a target nucleic acid is added to a reaction mixture that comprises the necessary amplification components, and a modified primer is formed.

In general, the modified primer comprises a detectable label, such as a fluorescent label, which is either incorporated by the enzyme or present on the original primer. As required, the unreacted primers are removed, in a variety of ways, as will be appreciated by those in the art and outlined herein. The hybridization complex is then disassociated, and the modified primer is detected and optionally quantitated by a microsphere array. In some cases, the newly modified primer serves as a target sequence for a secondary reaction, which then produces a number of amplified strands, which can be detected as outlined herein.

Accordingly, the present invention provides compositions and methods for detecting the presence or absence of target nucleic acid sequences in a sample. As will be appreciated by those in the art, the sample solution may comprise any number of things, including, but not limited to, bodily fluids (including, but not limited to, blood, urine, serum, lymph, saliva, anal and vaginal secretions, perspiration and semen, of virtually any organism, with mammalian samples being preferred and

human samples being particularly preferred); environmental samples (including, but not limited to, air, agricultural, water and soil samples); biological warfare agent samples; research samples; purified samples, such as purified genomic DNA, RNA, proteins, etc.; raw samples (bacteria, virus, genomic DNA, etc.; As will be appreciated by those in the art, virtually any experimental manipulation may have been done on the sample.

The present invention provides compositions and methods for detecting the presence or absence of target nucleic acid sequences in a sample. By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together. A nucleic acid of the present invention will generally contain phosphodiester bonds, although in some cases, as outlined below, nucleic acid analogs are included that may have alternate backbones, comprising, for example, phosphoramide (Beaucage et al., Tetrahedron 49(10):1925 (1993) and references therein; Letsinger, J. Org. Chem. 35:3800 (1970); Sprinzl et al., Eur. J. Biochem. 81:579 (1977); Letsinger et al., Nucl. Acids Res. 14:3487 (1986); Sawai et al, Chem. Lett. 805 (1984), Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); and Pauwels et al., Chemica Scripta 26:141 91986)), phosphorothioate (Mag et al., Nucleic Acids Res. 19:1437 (1991); and U.S. Patent No. 5,644,048), phosphorodithioate (Briu et al., J. Am. Chem. Soc. 111:2321 (1989), O-methylphophoroamidite linkages (see Eckstein, Oligonucleotides and Analogues: A Practical Approach, Oxford University Press), and peptide nucleic acid backbones and linkages (see Egholm, J. Am. Chem. Soc. 114:1895 (1992); Meier et al., Chem. Int. Ed. Engl. 31:1008 (1992); Nielsen, Nature, 365:566 (1993); Carlsson et al., Nature 380:207 (1996), all of which are incorporated by reference). Other analog nucleic acids include those with positive backbones (Denpcy et al., Proc. Natl. Acad. Sci. USA 92:6097 (1995); non-ionic backbones (U.S. Patent Nos. 5,386,023, 5,637,684, 5,602,240, 5,216,141 and 4,469,863; Kiedrowshi et al., Angew. Chem. Intl. Ed. English 30:423 (1991); Letsinger et al., J. Am. Chem. Soc. 110:4470 (1988); Letsinger et al., Nucleoside & Nucleotide 13:1597 (1994); Chapters 2 and 3, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook; Mesmaeker et al., Bioorganic & Medicinal Chem. Lett. 4:395 (1994); Jeffs et al., J. Biomolecular NMR 34:17 (1994); Tetrahedron Lett. 37:743 (1996)) and non-ribose backbones, including those described in U.S. Patent Nos. 5,235,033 and 5,034,506, and Chapters 6 and 7, ASC Symposium Series 580, "Carbohydrate Modifications in Antisense Research", Ed. Y.S. Sanghui and P. Dan Cook. Nucleic acids containing one or more carbocyclic sugars are also included within the definition of nucleic acids (see Jenkins et al., Chem. Soc. Rev. (1995) pp169-176). Several nucleic acid analogs are described in Rawls, C & E News June 2, 1997 page 35. All of these references are hereby expressly incorporated by reference. These modifications of the ribose-phosphate backbone may be done to facilitate the addition of labels, or to increase the stability and half-life of such molecules in physiological environments.

As will be appreciated by those in the art, all of these nucleic acid analogs may find use in the present invention. In addition, mixtures of naturally occurring nucleic acids and analogs can be made.

Alternatively, mixtures of different nucleic acid analogs, and mixtures of naturally occuring nucleic acids and analogs may be made.

Particularly preferred are peptide nucleic acids (PNA) which includes peptide nucleic acid analogs. These backbones are substantially non-ionic under neutral conditions, in contrast to the highly charged phosphodiester backbone of naturally occurring nucleic acids. This results in two advantages. First, the PNA backbone exhibits improved hybridization kinetics. PNAs have larger changes in the melting temperature (Tm) for mismatched versus perfectly matched basepairs. DNA and RNA typically exhibit a 2-4°C drop in Tm for an internal mismatch. With the non-ionic PNA backbone, the drop is closer to 7-9°C. This allows for better detection of mismatches. Similarly, due to their non-ionic nature, hybridization of the bases attached to these backbones is relatively insensitive to salt concentration.

The nucleic acids may be single stranded or double stranded, as specified, or contain portions of both double stranded or single stranded sequence. The nucleic acid may be DNA, both genomic and cDNA, RNA or a hybrid, where the nucleic acid contains any combination of deoxyribo- and ribo-nucleotides, and any combination of bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xathanine hypoxathanine, isocytosine, isoguanine, etc. A preferred embodiment utilizes isocytosine and isoguanine in nucleic acids designed to be complementary to other probes, rather than target sequences, as this reduces non-specific hybridization, as is generally described in U.S. Patent No. 5,681,702. As used herein, the term "nucleoside" includes nucleotides as well as nucleoside and nucleotide analogs, and modified nucleosides such as amino modified nucleosides. In addition, "nucleoside" includes non-naturally occuring analog structures. Thus for example the individual units of a peptide nucleic acid, each containing a base, are referred to herein as a nucleoside.

The compositions and methods of the invention are directed to the detection of target sequences. The term "target sequence" or "target nucleic acid" or grammatical equivalents herein means a nucleic acid sequence on a single strand of nucleic acid. The target sequence may be a portion of a gene, a regulatory sequence, genomic DNA, cDNA, RNA including mRNA and rRNA, or others. As is outlined herein, the target sequence may be a target sequence from a sample, or a secondary target such as a product of a reaction such as a detection sequence from an invasive cleavage reaction, a ligated probe from an OLA reaction, an extended probe from a PCR reaction, etc. Generally, as outlined herein, a target sequence from a sample is amplified to produce a secondary target that is detected, as outlined herein. Alternatively, an amplification step is done using a signal probe that is amplified, again producing a secondary target that is detected. The target sequence may be any length, with the understanding that longer sequences are more specific. As will be appreciated by those in the art, the complementary target sequence may take many forms. For example, it may be contained within a larger nucleic acid sequence, i.e. all or part of a gene or mRNA, a restriction fragment of a plasmid or

genomic DNA, among others. As is outlined more fully below, probes are made to hybridize to target sequences to determine the presence or absence of the target sequence in a sample. Generally speaking, this term will be understood by those skilled in the art. The target sequence may also be comprised of different target domains; for example, in "sandwich" type assays as outlined below, a first target domain of the sample target sequence may hybridize to a capture probe or a portion of capture extender probe, a second target domain may hybridize to a portion of an amplifier probe, a label probe, or a different capture or capture extender probe, etc. In addition, the target domains may be adjacent (i.e. contiguous) or separated. For example, when LCR techniques are used, a first primer may hybridize to a first target domain and a second primer may hybridize to a second target domain; either the domains are adjacent, or they may be separated by one or more nucleotides, coupled with the use of a polymerase and dNTPs, as is more fully outlined below. The terms "first" and "second" are not meant to confer an orientation of the sequences with respect to the 5'-3' orientation of the target sequence. For example, assuming a 5'-3' orientation of the complementary target sequence, the first target domain may be located either 5' to the second domain, or 3' to the second domain.

If required, the target sequence is prepared using known techniques. For example, the sample may be treated to lyse the cells, using known lysis buffers, sonication, electroporation, etc., with purification occuring as needed, as will be appreciated by those in the art. In addition, the reactions outlined herein may be accomplished in a variety of ways, as will be appreciated by those in the art. Components of the reaction may be added simultaneously, or sequentially, in any order, with preferred embodiments outlined below. In addition, the reaction may include a variety of other reagents which may be included in the assays. These include reagents like salts, buffers, neutral proteins, e.g. albumin, detergents, etc., which may be used to facilitate optimal hybridization and detection, and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used, depending on the sample preparation methods and purity of the target.

In addition, in most embodiments, double stranded target nucleic acids are denatured to render them single stranded so as to permit hybridization of the primers and other probes of the invention. A preferred embodiment utilizes a thermal step, generally by raising the temperature of the reaction to about 95°C, although pH changes and other techniques may also be used.

A primer nucleic acid is then contacted to the target sequence to form a hybridization complex. By "primer nucleic acid" herein is meant a probe nucleic acid that will hybridize to some portion, i.e. a domain, of the target sequence. Probes of the present invention are designed to be complementary to a target sequence (either the target sequence of the sample or to other probe sequences, as is described below), such that hybridization of the target sequence and the probes of the present invention occurs. As outlined below, this complementarity need not be perfect; there may be any number of base pair mismatches which will interfere with hybridization between the target sequence

and the single stranded nucleic acids of the present invention. However, if the number of mutations is so great that no hybridization can occur under even the least stringent of hybridization conditions, the sequence is not a complementary target sequence. Thus, by "substantially complementary" herein is meant that the probes are sufficiently complementary to the target sequences to hybridize under normal reaction conditions.

A variety of hybridization conditions may be used in the present invention, including high, moderate and low stringency conditions; see for example Maniatis et al., Molecular Cloning: A Laboratory Manual, 2d Edition, 1989, and Short Protocols in Molecular Biology, ed. Ausubel, et al, hereby incorporated by reference. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, stringent conditions are selected to be about 5-10°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The Tm is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at Tm, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (e.g. 10 to 50 nucleotides) and at least about 60°C for long probes (e.g. greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of helix destabilizing agents such as formamide. The hybridization conditions may also vary when a non-ionic backbone, i.e. PNA is used, as is known in the art. In addition, cross-linking agents may be added after target binding to cross-link, i.e. covalently attach, the two strands of the hybridization complex.

Thus, the assays are generally run under stringency conditions which allows formation of the hybridization complex only in the presence of target. Stringency can be controlled by altering a step parameter that is a thermodynamic variable, including, but not limited to, temperature, formamide concentration, salt concentration, chaotropic salt concentration, pH, organic solvent concentration, etc.

These parameters may also be used to control non-specific binding, as is generally outlined in U.S. Patent No. 5,681,697. Thus it may be desirable to perform certain steps at higher stringency conditions to reduce non-specific binding.

The size of the primer nucleic acid may vary, as will be appreciated by those in the art, in general varying from 5 to 500 nucleotides in length, with primers of between 10 and 100 being preferred, between 15 and 50 being particularly preferred, and from 10 to 35 being especially preferred,

depending on the use and amplification technique.

In addition, the different amplification techniques may have further requirements of the primers, as is more fully described below.

Once the hybridization complex between the primer and the target sequence has been formed, an enzyme, sometimes termed an "amplification enzyme", is used to modify the primer. As for all the methods outlined herein, the enzymes may be added at any point during the assay, either prior to, during, or after the addition of the primers. The identity of the enzyme will depend on the amplification technique used, as is more fully outlined below. Similarly, the modification will depend on the amplification technique, as outlined below.

Once the enzyme has modified the primer to form a modified primer, the hybridization complex is disassociated. In one aspect, dissociation is by modification of the assay conditions. In another aspect, the modified primer no longer hybridizes to the target nucleic acid and dissociates. Either one or both of these aspects can be employed in signal and target amplification reactions as described below. Generally, the amplification steps are repeated for a period of time to allow a number of cycles, depending on the number of copies of the original target sequence and the sensitivity of detection, with cycles ranging from 1 to thousands, with from 10 to 100 cycles being preferred and from 20 to 50 cycles being especially preferred.

After a suitable time of amplification, unreacted primers are removed, in a variety of ways, as will be appreciated by those in the art and described below, and the hybridization complex is disassociated. In general, the modified primer comprises a detectable label, such as a fluorescent label, which is either incorporated by the enzyme or present on the original primer, and the modified primer is added to a microsphere array such is generally described in U.S.S.N.s 09/189,543; 08/944,850; 09/033,462; 09/287,573; 09/151,877; 09/187,289 and 09/256,943; and PCT applications US98/09163 and US99/14387; US98/21193; US99/04473 and US98/05025, all of which are hereby incorporated by reference. The microsphere array comprises subpopulations of microspheres that comprise capture probes that will hybridize to the modified primers. Detection proceeds via detection of the label as an indication of the presence, absence or amount of the target sequence, as is more fully outlined below.

TARGET AMPLIFICATION

In a preferred embodiment, the amplification is target amplification. Target amplification involves the amplification (replication) of the target sequence to be detected, such that the number of copies of the target sequence is increased. Suitable target amplification techniques include, but are not limited to, the polymerase chain reaction (PCR), strand displacement amplification (SDA), transcription mediated amplification (TMA) and nucleic acid sequence based amplification (NASBA).

POLYMERASE CHAIN REACTION AMPLIFICATION

In a preferred embodiment, the target amplification technique is PCR. The polymerase chain reaction (PCR) is widely used and described, and involves the use of primer extension combined with thermal cycling to amplify a target sequence; see U.S. Patent Nos. 4,683,195 and 4,683,202, and PCR Essential Data, J. W. Wiley & sons, Ed. C.R. Newton, 1995, all of which are incorporated by reference. In addition, there are a number of variations of PCR which also find use in the invention, including "quantitative competitive PCR" or "QC-PCR", "arbitrarily primed PCR" or "AP-PCR" , "immuno-PCR", "Alu-PCR", "PCR single strand conformational polymorphism" or "PCR-SSCP", "reverse transcriptase PCR" or "RT-PCR", "biotin capture PCR", "vectorette PCR", "panhandle PCR", and "PCR select cDNA subtraction", "allele-specific PCR", among others.

In general, PCR may be briefly described as follows. A double stranded target nucleic acid is denatured, generally by raising the temperature, and then cooled in the presence of an excess of a PCR primer, which then hybridizes to the first target strand. A DNA polymerase then acts to extend the primer with dNTPs, resulting in the synthesis of a new strand forming a hybridization complex. The sample is then heated again, to disassociate the hybridization complex, and the process is repeated. By using a second PCR primer for the complementary target strand, rapid and exponential amplification occurs. Thus PCR steps are denaturation, annealing and extension. The particulars of PCR are well known, and include the use of a thermostable polymerase such as Taq I polymerase and thermal cycling.

Accordingly, the PCR reaction requires at least one PCR primer, a polymerase, and a set of dNTPs. As outlined herein, the primers may comprise the label, or one or more of the dNTPs may comprise a label.

In general, as is more fully outlined below, the capture probes on the beads of the array are designed to be substantially complementary to the extended part of the primer; that is, unextended primers will not bind to the capture probes. Alternatively, as further described below, unreacted probes may be removed prior to addition to the array.

STRAND DISPLACEMENT AMPLIFICATION (SDA)

In a preferred embodiment, the target amplification technique is SDA. Strand displacement amplification (SDA) is generally described in Walker et al., in Molecular Methods for Virus Detection, Academic Press, Inc., 1995, and U.S. Patent Nos. 5,455,166 and 5,130,238, all of which are hereby expressly incorporated by reference in their entirety.

In general, SDA may be described as follows. A single stranded target nucleic acid, usually a DNA target sequence, is contacted with an SDA primer. An "SDA primer" generally has a length of 25-100 nucleotides, with SDA primers of approximately 35 nucleotides being preferred. An SDA primer is

substantially complementary to a region at the 3' end of the target sequence, and the primer has a sequence at its 5' end (outside of the region that is complementary to the target) that is a recognition sequence for a restriction endonuclease, sometimes referred to herein as a "nicking enzyme" or a "nicking endonuclease", as outlined below. The SDA primer then hybridizes to the target sequence. The SDA reaction mixture also contains a polymerase (an "SDA polymerase", as outlined below) and a mixture of all four deoxynucleoside-triphosphates (also called deoxynucleotides or dNTPs, i.e. dATP, dTTP, dCTP and dGTP), at least one species of which is a substituted or modified dNTP; thus, the SDA primer is modified, i.e. extended, to form a modified primer, sometimes referred to herein as a "newly synthesized strand". The substituted dNTP is modified such that it will inhibit cleavage in the strand containing the substituted dNTP but will not inhibit cleavage on the other strand. Examples of suitable substituted dNTPs include, but are not limited, 2'deoxyadenosine 5'-O-(1-thiotriphosphate), 5-methyldeoxycytidine 5'-triphosphate, 2'-deoxyuridine 5'-triphosphate, adn 7-deaza-2'-deoxyguanosine 5'-triphosphate. In addition, the substitution of the dNTP may occur after incorporation into a newly synthesized strand; for example, a methylase may be used to add methyl groups to the synthesized strand. In addition, if all the nucleotides are substituted, the polymerase may have 5'→3' exonuclease activity. However, if less than all the nucleotides are substituted, the polymerase preferably lacks 5'→3' exonuclease activity.

As will be appreciated by those in the art, the recognition site/endonuclease pair can be any of a wide variety of known combinations. The endonuclease is chosen to cleave a strand either at the recognition site, or either 3' or 5' to it, without cleaving the complementary sequence, either because the enzyme only cleaves one strand or because of the incorporation of the substituted nucleotides. Suitable recognition site/endonuclease pairs are well known in the art; suitable endonucleases include, but are not limited to, HincII, HindII, AvaI, Fnu4HI, TthIIII, NcII, BstXI, BamHI, etc. A chart depicting suitable enzymes, and their corresponding recognition sites and the modified dNTP to use is found in U.S. Patent No. 5,455,166, hereby expressly incorporated by reference.

Once nicked, a polymerase (an "SDA polymerase") is used to extend the newly nicked strand, 5'→3', thereby creating another newly synthesized strand. The polymerase chosen should be able to intiate 5'→3' polymerization at a nick site, should also displace the polymerized strand downstream from the nick, and should lack 5'→3' exonuclease activity (this may be additionally accomplished by the addition of a blocking agent). Thus, suitable polymerases in SDA include, but are not limited to, the Klenow fragment of DNA polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase.

Accordingly, the SDA reaction requires, in no particular order, an SDA primer, an SDA polymerase, a nicking endonuclease, and dNTPs, at least one species of which is modified. Again, as outlined above for PCR, preferred embodiments utilize capture probes complementary to the newly synthesized portion of the primer, rather than the primer region, to allow unextended primers to be

removed.

In general, SDA does not require thermocycling. The temperature of the reaction is generally set to be high enough to prevent non-specific hybridization but low enough to allow specific hybridization; this is generally from about 37°C to about 42°C, depending on the enzymes.

In a preferred embodiment, as for most of the amplification techniques described herein, a second amplification reaction can be done using the complementary target sequence, resulting in a substantial increase in amplification during a set period of time. That is, a second primer nucleic acid is hybridized to a second target sequence, that is substantially complementary to the first target sequence, to form a second hybridization complex. The addition of the enzyme, followed by disassociation of the second hybridization complex, results in the generation of a number of newly synthesized second strands.

NUCLEIC ACID SEQUENCE BASED AMPLIFICATION (NASBA) AND TRANSCRIPTION MEDIATED AMPLIFICATION (TMA)

In a preferred embodiment, the target amplification technique is nucleic acid sequence based amplification (NASBA). NASBA is generally described in U.S. Patent No. 5,409,818; Sooknanan et al., Nucleic Acid Sequence-Based Amplification, Ch. 12 (pp. 261-285) of Molecular Methods for Virus Detection, Academic Press, 1995; and "Profiting from Gene-based Diagnostics", CTB International Publishing Inc., N.J., 1996, all of which are incorporated by reference. NASBA is very similar to both TMA and QBR. Transcription mediated amplification (TMA) is generally described in U.S. Patent Nos. 5,399,491, 5,888,779, 5,705,365, 5,710,029, all of which are incorporated by reference. The main difference between NASBA and TMA is that NASBA utilizes the addition of RNAse H to effect RNA degradation, and TMA relies on inherent RNAse H activity of the reverse transcriptase.

In general, these techniques may be described as follows. A single stranded target nucleic acid, usually an RNA target sequence (sometimes referred to herein as "the first target sequence" or "the first template"), is contacted with a first primer, generally referred to herein as a "NASBA primer" (although "TMA primer" is also suitable). Starting with a DNA target sequence is described below. These primers generally have a length of 25-100 nucleotides, with NASBA primers of approximately 50-75 nucleotides being preferred. The first primer is preferably a DNA primer that has at its 3' end a sequence that is substantially complementary to the 3' end of the first template. The first primer also has an RNA polymerase promoter at its 5' end (or its complement (antisense), depending on the configuration of the system). The first primer is then hybridized to the first template to form a first hybridization complex. The reaction mixture also includes a reverse transcriptase enzyme (an "NASBA reverse transcriptase") and a mixture of the four dNTPs, such that the first NASBA primer is modified, i.e. extended, to form a modified first primer, comprising a hybridization complex of RNA (the first template) and DNA (the newly synthesized strand).

By "reverse transcriptase" or "RNA-directed DNA polymerase" herein is meant an enzyme capable of synthesizing DNA from a DNA primer and an RNA template. Suitable RNA-directed DNA polymerases include, but are not limited to, avian myloblastosis virus reverse transcriptase ("AMV RT") and the Moloney murine leukemia virus RT. When the amplification reaction is TMA, the reverse transcriptase enzyme further comprises a RNA degrading activity as outlined below.

In addition to the components listed above, the NASBA reaction also includes an RNA degrading enzyme, also sometimes referred to herein as a ribonuclease, that will hydrolyze RNA of an RNA:DNA hybrid without hydrolyzing single- or double-stranded RNA or DNA. Suitable ribonucleases include, but are not limited to, RNase H from *E. coli* and calf thymus.

The ribonuclease activity degrades the first RNA template in the hybridization complex, resulting in a disassociation of the hybridization complex leaving a first single stranded newly synthesized DNA strand, sometimes referred to herein as "the second template".

In addition, the NASBA reaction also includes a second NASBA primer, generally comprising DNA (although as for all the probes herein, including primers, nucleic acid analogs may also be used). This second NASBA primer has a sequence at its 3' end that is substantially complementary to the 3' end of the second template, and also contains an antisense sequence for a functional promoter and the antisense sequence of a transcription initiation site. Thus, this primer sequence, when used as a template for synthesis of the third DNA template, contains sufficient information to allow specific and efficient binding of an RNA polymerase and initiation of transcription at the desired site. Preferred embodiments utilizes the antisense promoter and transcription initiation site are that of the T7 RNA polymerase, although other RNA polymerase promoters and initiation sites can be used as well, as outlined below.

The second primer hybridizes to the second template, and a DNA polymerase, also termed a "DNA-directed DNA polymerase", also present in the reaction, synthesizes a third template (a second newly synthesized DNA strand), resulting in second hybridization complex comprising two newly synthesized DNA strands.

Finally, the inclusion of an RNA polymerase and the required four ribonucleoside triphosphates (ribonucleotides or NTPs) results in the synthesis of an RNA strand (a third newly synthesized strand that is essentially the same as the first template). The RNA polymerase, sometimes referred to herein as a "DNA-directed RNA polymerase", recognizes the promoter and specifically initiates RNA synthesis at the initiation site. In addition, the RNA polymerase preferably synthesizes several copies of RNA per DNA duplex. Preferred RNA polymerases include, but are not limited to, T7 RNA polymerase, and other bacteriophage RNA polymerases including those of phage T3, phage ϕII, Salmonella phage sp6, or Pseudomonase phage gh-1.

In some embodiments, TMA and NASBA are used with starting DNA target sequences. In this embodiment, it is necessary to utilize the first primer comprising the RNA polymerase promoter and a DNA polymerase enzyme to generate a double stranded DNA hybrid with the newly synthesized strand comprising the promoter sequence. The hybrid is then denatured and the second primer added.

Accordingly, the NASBA reaction requires, in no particular order, a first NASBA primer, a second NASBA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase, a DNA polymerase, an RNA degrading enzyme, NTPs and dNTPs, in addition to the detection components outlined below.

These components result in a single starting RNA template generating a single DNA duplex; however, since this DNA duplex results in the creation of multiple RNA strands, which can then be used to initiate the reaction again, amplification proceeds rapidly.

Accordingly, the TMA reaction requires, in no particular order, a first TMA primer, a second TMA primer comprising an antisense sequence of an RNA polymerase promoter, an RNA polymerase that recognizes the promoter, a reverse transcriptase with RNA degrading activity, a DNA polymerase, NTPs and dNTPs, in addition to the detection components outlined below.

These components result in a single starting RNA template generating a single DNA duplex; however, since this DNA duplex results in the creation of multiple RNA strands, which can then be used to initiate the reaction again, amplification proceeds rapidly.

As outlined herein, the detection of the newly synthesized strands can proceed in several ways. Direct detection can be done when the newly synthesized strands comprise detectable labels, either by incorporation into the primers or by incorporation of modified labelled nucleotides into the growing strand. Alternatively, as is more fully outlined below, indirect detection of unlabelled strands (which now serve as "targets" in the detection mode) can occur using a variety of sandwich assay configurations. As will be appreciated by those in the art, any of the newly synthesized strands can serve as the "target" for form an assay complex on a surface with a capture probe. In NASBA and TMA, it is preferable to utilize the newly formed RNA strands as the target, as this is where significant amplification occurs.

In this way, a number of secondary target molecules are made. As is more fully outlined below, these reactions (that is, the products of these reactions) can be detected in a number of ways.

SIGNAL AMPLIFICATION TECHNIQUES

In a preferred embodiment, the amplification technique is signal amplification. Signal amplification

involves the use of limited number of target molecules as templates to either generate multiple signalling probes or allow the use of multiple signalling probes. Signal amplification strategies include LCR, CPT, QβR, invasive cleavage technology, and the use of amplification probes in sandwich assays.

SINGLE BASE EXTENSION (SBE)

In a preferred embodiment, single base extension (SBE; sometimes referred to as "minisequencing") is used for amplification. It should also be noted that SBE finds use in genotyping, as is described in co-pending application entitled "SEQUENCE DETERMINATION OF NUCLEIC ACIDS USING ARRAYS WITH MICROSPHERES" filed on October 22, 1999 as U.S. Serial No. 09/425,633. Briefly, SBE is a technique that utilizes an extension primer that hybridizes to the target nucleic acid. A polymerase (generally a DNA polymerase) is used to extend the 3' end of the primer with a nucleotide analog labeled a detection label as described herein. Based on the fidelity of the enzyme, a nucleotide is only incorporated into the extension primer if it is complementary to the adjacent base in the target strand. Generally, the nucleotide is derivatized such that no further extensions can occur, so only a single nucleotide is added. However, for amplification reactions, this may not be necessary. Once the labeled nucleotide is added, detection of the label proceeds as outlined herein. See generally Sylvanen et al., Genomics 8:684-692 (1990); U.S. Patent Nos. 5,846,710 and 5,888,819; Pastinen et al., Genomics Res. 7(6):606-614 (1997); all of which are expressly incorporated herein by reference.

The reaction is initiated by introducing the assay complex comprising the target sequence (i.e. the array) to a solution comprising a first nucleotide, frequently an nucleotide analog. By "nucleotide analog" in this context herein is meant a deoxynucleoside-triphosphate (also called deoxynucleotides or dNTPs, i.e. dATP, dTTP, dCTP and dGTP), that is further derivatized to be chain terminating. As will be appreciated by those in the art, any number of nucleotide analogs may be used, as long as a polymerase enzyme will still incorporate the nucleotide at the interrogation position. Preferred embodiments utilize dideoxy-triphosphate nucleotides (ddNTPs). Generally, a set of nucleotides comprising ddATP, ddCTP, ddGTP and ddTTP is used, at least one of which includes a label, and preferably all four. For amplification rather than genotyping reactions, the labels may all be the same; alternatively, different labels may be used.

In a preferred embodiment, the nucleotide analogs comprise a detectable label, which can be either a primary or secondary detectable label. Preferred primary labels are those outlined above. However, the enzymatic incorporation of nucleotides comprising fluorophores is poor under many conditions; accordingly, preferred embodiments utilize secondary detectable labels. In addition, as outlined below, the use of secondary labels may also facilitate the removal of unextended probes.

In addition to a first nucleotide, the solution also comprises an extension enzyme, generally a DNA polymerase. Suitable DNA polymerases include, but are not limited to, the Klenow fragment of DNA

polymerase I, SEQUENASE 1.0 and SEQUENASE 2.0 (U.S. Biochemical), T5 DNA polymerase and Phi29 DNA polymerase. If the NTP is complementary to the base of the detection position of the target sequence, which is adjacent to the extension primer, the extension enzyme will add it to the extension primer. Thus, the extension primer is modified, i.e. extended, to form a modified primer, sometimes referred to herein as a "newly synthesized strand".

A limitation of this method is that unless the target nucleic acid is in sufficient concentration, the amount of unextended primer in the reaction greatly exceeds the resultant extended-labeled primer. The excess of unextended primer competes with the detection of the labeled primer in the assays described herein. Accordingly, when SBE is used, preferred embodiments utilize methods for the removal of unextended primers as outlined herein.

One method to overcome this limitation is thermocycling minisequencing in which repeated cycles of annealing, primer extension, and heat denaturation using a thermocycler and thermo-stable polymerase allows the amplification of the extension probe which results in the accumulation of extended primers. For example, if the original unextended primer to target nucleic acid concentration is 100:1 and 100 thermocycles and extensions are performed, a majority of the primer will be extended.

As will be appreciated by those in the art, the configuration of the SBE system can take on several forms. As for the LCR reaction described below, the reaction may be done in solution, and then the newly synthesized strands, with the base-specific detectable labels, can be detected. For example, they can be directly hybridized to capture probes that are complementary to the extension primers, and the presence of the label is then detected.

Alternatively, the SBE reaction can occur on a surface. For example, a target nucleic acid may be captured using a first capture probe that hybridizes to a first target domain of the target, and the reaction can proceed at a second target domain. The extended labeled primers are then bound to a second capture probe and detected.

Thus, the SBE reaction requires, in no particular order, an extension primer, a polymerase and dNTPs, at least one of which is labeled.

OLIGONUCLEOTIDE LIGATION AMPLIFICATION (OLA)

In a preferred embodiment, the signal amplification technique is OLA. OLA, which is referred to as the ligation chain reaction (LCR) when two-stranded substrates are used, involves the ligation of two smaller probes into a single long probe, using the target sequence as the template. In LCR, the ligated probe product becomes the predominant template as the reaction progresses. The method can be run in two different ways; in a first embodiment, only one strand of a target sequence is used as a

template for ligation; alternatively, both strands may be used. See generally U.S. Patent Nos. 5,185,243, 5,679,524 and 5,573,907; EP 0 320 308 B1; EP 0 336 731 B1; EP 0 439 182 B1; WO 90/01069; WO 89/12696; WO 97/31256; and WO 89/09835, and U.S.S.N.s 60/078,102 and 60/073,011, all of which are incorporated by reference.

In a preferred embodiment, the single-stranded target sequence comprises a first target domain and a second target domain, which are adjacent and contiguous. A first OLA primer and a second OLA primer nucleic acids are added, that are substantially complementary to their respective target domain and thus will hybridize to the target domains. These target domains may be directly adjacent, i.e. contiguous, or separated by a number of nucleotides. If they are non-contiguous, nucleotides are added along with means to join nucleotides, such as a polymerase, that will add the nucleotides to one of the primers. The two OLA primers are then covalently attached, for example using a ligase enzyme such as is known in the art, to form a modified primer. This forms a first hybridization complex comprising the ligated probe and the target sequence. This hybridization complex is then denatured (disassociated), and the process is repeated to generate a pool of ligated probes.

In a preferred embodiment, OLA is done for two strands of a double-stranded target sequence. The target sequence is denatured, and two sets of probes are added: one set as outlined above for one strand of the target, and a separate set (i.e. third and fourth primer probe nucleic acids) for the other strand of the target. In a preferred embodiment, the first and third probes will hybridize, and the second and fourth probes will hybridize, such that amplification can occur. That is, when the first and second probes have been attached, the ligated probe can now be used as a template, in addition to the second target sequence, for the attachment of the third and fourth probes. Similarly, the ligated third and fourth probes will serve as a template for the attachment of the first and second probes, in addition to the first target strand. In this way, an exponential, rather than just a linear, amplification can occur.

As will be appreciated by those in the art, the ligation product can be detected in a variety of ways. In a preferred embodiment, the ligation reaction is run in solution. In this embodiment, only one of the primers carries a detectable label, e.g. the first ligation probe, and the capture probe on the bead is substantially complementary to the other probe, e.g. the second ligation probe. In this way, unextended labeld ligation primers will not interfere with the assay. That is, in a preferred embodiment, the ligation product is detected by solid-phase oligonucleotide probes. The solid-phase probes are preferably complementary to at least a portion of the ligation product. In a preferred embodiment, the solid-phase probe is complementary to the 5' detection oligonucleotide portion of the ligation product. This substantially reduces or eliminates false signal generated by the optically-labeled 3' primers. Preferably, detection is accomplished by removing the unligated 5' detection oligonucleotide from the reaction before application to a capture probe. In one embodiment, the unligated 5' detection oligonucleotides are removed by digesting 3' non-protected oligonucleotides with a 3' exonuclease,

such as, exonuclease I. The ligation products are protected from exo I digestion by including, for example, 4-phosphorothioate residues at their 3' terminus, thereby, rendering them resistant to exonuclease digestion. The unligated detection oligonucleotides are not protected and are digested.

Alternatively, the target nucleic acid is immobilized on a solid-phase surface. The ligation assay is performed and unligated oligonucleotides are removed by washing under appropriate stringency to remove unligated oligonucleotides. The ligated oligonucleotides are eluted from the target nucleic acid using denaturing conditions, such as, 0.1 N NaOH, and detected as described herein.

Again, as outlined above, the detection of the LCR reaction can also occur directly, in the case where one or both of the primers comprises at least one detectable label, or indirectly, using sandwich assays, through the use of additional probes; that is, the ligated probes can serve as target sequences, and detection may utilize amplification probes, capture probes, capture extender probes, label probes, and label extender probes, etc.

ROLLING-CIRCLE AMPLIFICATION (RCA)

In a preferred embodiment the signal amplification technique is RCA. Rolling-circle amplification is generally described in Baner *et al.* (1998) *Nuc. Acids Res.* 26:5073-5078; Barany, F. (1991) *Proc. Natl. Acad. Sci. USA* 88:189-193; and Lizardi *et al.* (1998) *Nat. Genet.* 19:225-232, all of which are incorporated by reference in their entirety.

In general, RCA may be described in two ways. First, as is outlined in more detail below, a single probe is hybridized with a target nucleic acid. Each terminus of the probe hybridizes adjacently on the target nucleic acid and the OLA assay as described above occurs. When ligated, the probe is circularized while hybridized to the target nucleic acid. Addition of a polymerase results in extension of the circular probe. However, since the probe has no terminus, the polymerase continues to extend the probe repeatedly. Thus results in amplification of the circular probe.

A second alternative approach involves OLA followed by RCA. In this embodiment, an immobilized primer is contacted with a target nucleic acid. Complementary sequences will hybridize with each other resulting in an immobilized duplex. A second primer is contacted with the target nucleic acid. The second primer hybridizes to the target nucleic acid adjacent to the first primer. An OLA assay is performed as described above. Ligation only occurs if the primer are complementary to the target nucleic acid. When a mismatch occurs, particularly at one of the nucleotides to be ligated, ligation will not occur. Following ligation of the oligonucleotides, the ligated, immobilized, oligonucleotide is then hybridized with an RCA probe. This is a circular probe that is designed to specifically hybridize with the ligated oligonucleotide and will only hybridize with an oligonucleotide that has undergone ligation. RCA is then performed as is outlined in more detail below.

Accordingly, in an preferred embodiment, a single oligonucleotide is used both for OLA and as the circular template for RCA (referred to herein as a "padlock probe" or a "RCA probe"). That is, each terminus of the oligonucleotide contains sequence complementary to the target nucleic acid and functions as an OLA primer as described above. That is, the first end of the RCA probe is substantially complementary to a first target domain, and the second end of the RCA probe is substantially complementary to a second target domain, adjacent to the first domain. Hybridization of the oligonucleotide to the target nucleic acid results in the formation of a hybridization complex. Ligation of the "primers" (which are the discrete ends of a single oligonucleotide) results in the formation of a modified hybridization complex containing a circular probe i.e. an RCA template complex. That is, the oligonucleotide is circularized while still hybridized with the target nucleic acid. This serves as a circular template for RCA. Addition of a polymerase to the RCA template complex results in the formation of an amplified product nucleic acid. Following RCA, the amplified product nucleic acid is detected (Figure 6). This can be accomplished in a variety of ways; for example, the polymerase may incorporate labelled nucleotides, or alternatively, a label probe is used that is substantially complementary to a portion of the RCA probe and comprises at least one label is used.

The polymerase can be any polymerase, but is preferably one lacking 3' exonuclease activity (3' exo$^-$). Examples of suitable polymerase include but are not limited to exonuclease minus DNA Polymerase I large (Klenow) Fragment, Phi29 DNA polymerase, Taq DNA Polymerase and the like. In addition, in some embodiments, a polymerase that will replicate single-stranded DNA (i.e. without a primer forming a double stranded section) can be used.

In a preferred embodiment, the RCA probe contains an adapter sequence as outlined herein, with adapter capture probes on the array, for example on a microsphere when microsphere arrays are being used. Alternatively, unique portions of the RCA probes, for example all or part of the sequence corresponding to the target sequence, can be used to bind to a capture probe.

In a preferred embodiment, the padlock probe contains a restriction site. The restriction endonuclease site allows for cleavage of the long concatamers that are typically the result of RCA into smaller individual units that hybridize either more efficiently or faster to surface bound capture probes. Thus, following RCA, the product nucleic acid is contacted with the appropriate restriction endonuclease. This results in cleavage of the product nucleic acid into smaller fragments. The fragments are then hybridized with the capture probe that is immobilized resulting in a concentration of product fragments onto the microsphere. Again, as outlined herein, these fragments can be detected in one of two ways: either labelled nucleotides are incorporated during the replication step, or an additional label probe is added.

Thus, in a preferred embodiment, the padlock probe comprises a label sequence; i.e. a sequence that can be used to bind label probes and is substantially complementary to a label probe. In one

embodiment, it is possible to use the same label sequence and label probe for all padlock probes on an array; alternatively, each padlock probe can have a different label sequence.

The padlock probe also contains a priming site for priming the RCA reaction. That is, each padlock probe comprises a sequence to which a primer nucleic acid hybridizes forming a template for the polymerase. The primer can be found in any portion of the circular probe. In a preferred embodiment, the primer is located at a discrete site in the probe. In this embodiment, the primer site in each distinct padlock probe is identical, although this is not required. Advantages of using primer sites with identical sequences include the ability to use only a single primer oligonucleotide to prime the RCA assay with a plurality of different hybridization complexes. That is, the padlock probe hybridizes uniquely to the target nucleic acid to which it is designed. A single primer hybridizes to all of the unique hybridization complexes forming a priming site for the polymerase. RCA then proceeds from an identical locus within each unique padlock probe of the hybridization complexes.

In an alternative embodiment, the primer site can overlap, encompass, or reside within any of the above-described elements of the padlock probe. That is, the primer can be found, for example, overlapping or within the restriction site or the identifier sequence. In this embodiment, it is necessary that the primer nucleic acid is designed to base pair with the chosen primer site.

Thus, the padlock probe of the invention contains at each terminus, sequences corresponding to OLA primers. The intervening sequence of the padlock probe contain in no particular order, an adapter sequence and a restriction endonuclease site. In addition, the padlock probe contains a RCA priming site.

Thus, in a preferred embodiment the OLA/RCA is performed in solution followed by restriction endonuclease cleavage of the RCA product. The cleaved product is then applied to an array comprising beads, each bead comprising a probe complementary to the adapter sequence located in the padlock probe. The amplified adapter sequence correlates with a particular target nucleic acid. Thus the incorporation of an endonuclease site allows the generation of short, easily hybridizable sequences. Furthermore, the unique adapter sequence in each rolling circle padlock probe sequence allows diverse sets of nucleic acid sequences to be analyzed in parallel on an array, since each sequence is resolved on the basis of hybridization specificity.

In an alternative OLA/RCA method, one of the OLA primers is immobilized on the microsphere; the second primer is added in solution. Both primers hybridize with the target nucleic acid forming a hybridization complex as described above for the OLA assay.

As described herein, the microsphere is distributed on an array. In a preferred embodiment, a plurality of microspheres each with a unique OLA primer is distributed on the array.

Following the OLA assay, and either before, after or concurrently with distribution of the beads on the array, a segment of circular DNA is hybridized to the bead-based ligated oligonucleotide forming a modified hybridization complex. Addition of an appropriate polymerase (3' exo$^-$), as is known in the art, and corresponding reaction buffer to the array leads to amplification of the circular DNA. Since there is no terminus to the circular DNA, the polymerase continues to travel around the circular template generating extension product until it detaches from the template. Thus, a polymerase with high processivity can create several hundred or thousand copies of the circular template with all the copies linked in one contiguous strand.

Again, these copies are subsequently detected by one of two methods; either hybridizing a labeled oligo complementary to the circular target or via the incorporation of labeled nucleotides in the amplification reaction. The label is detected using conventional label detection methods as described herein.

In one embodiment, when the circular DNA contains sequences complementary to the ligated oligonucleotide it is preferable to remove the target DNA prior to contacting the ligated oligonucleotide with the circular DNA (See Fig 7). This is done by denaturing the double-stranded DNA by methods known in the art. In an alternative embodiment, the double stranded DNA is not denatured prior to contacting the circular DNA.

In an alternative embodiment, when the circular DNA contains sequences complementary to the target nucleic acid, it is preferable that the circular DNA is complementary at a site distinct from the site bound to the ligated oligonucleotide. In this embodiment it is preferred that the duplex between the ligated oligonucleotide and target nucleic acid is not denatured or disrupted prior to the addition of the circular DNA so that the target DNA remains immobilized to the bead.

Hybridization and washing conditions are well known in the art; various degrees of stringency can be used. In some embodiments it is not necessary to use stringent hybridization or washing conditions as only microspheres containing the ligated probes will effectively hybridize with the circular DNA; microspheres bound to DNA that did not undergo ligation (those without the appropriate target nucleic acid) will not hybridize as strongly with the circular DNA as those primers that were ligated. Thus, hybridization and/or washing conditions are used that discriminate between binding of the circular DNA to the ligated primer and the unligated primer.

Alternatively, when the circular probe is designed to hybridize to the target nucleic acid at a site distinct from the site bound to the ligated oligonucleotide, hybridization and washing conditions are used to remove or dissociate the target nucleic acid from unligated oligonucleotides while target nucleic acid hybridizing with the ligated oligonucleotides will remain bound to the beads. In this embodiment, the circular probe only hybridizes to the target nucleic acid when the target nucleic acid

is hybridized with a ligated oligonucleotide that is immobilized on a bead.

As is well known in the art, an appropriate polymerase (3' exo$^-$) is added to the array. The polymerase extends the sequence of a single-stranded DNA using double-stranded DNA as a primer site. In one embodiment, the circular DNA that has hybridized with the appropriate OLA reaction product serves as the primer for the polymerase. In the presence of an appropriate reaction buffer as is known in the art, the polymerase will extend the sequence of the primer using the single-stranded circular DNA as a template. As there is no terminus of the circular DNA, the polymerase will continue to extend the sequence of the circular DNA. In an alternative embodiment, the RCA probe comprises a discrete primer site located within the circular probe. Hybridization of primer nucleic acids to this primer site forms the polymerase template allowing RCA to proceed.

In a preferred embodiment, the polymerase creates more than 100 copies of the circular DNA. In more preferred embodiments the polymerase creates more than 1000 copies of the circular DNA; while in a most preferred embodiment the polymerase creates more than 10,000 copies or more than 50,000 copies of the template.

The amplified circular DNA sequence is then detected by methods known in the art and as described herein. Detection is accomplished by hybridizing with a labeled probe. The probe is labeled directly or indirectly. Alternatively, labeled nucleotides are incorporated into the amplified circular DNA product. The nucleotides can be labeled directly, or indirectly as is further described herein.

The RCA as described herein finds use in allowing highly specific and highly sensitive detection of nucleic acid target sequences. In particular, the method finds use in improving the multiplexing ability of DNA arrays and eliminating costly sample or target preparation. As an example, a substantial savings in cost can be realized by directly analyzing genomic DNA on an array, rather than employing an intermediate PCR amplification step. The method finds use in examining genomic DNA and other samples including mRNA.

In addition the RCA finds use in allowing rolling circle amplification products to be easily detected by hybridization to probes in a solid-phase format (e.g. an array of beads). An additional advantage of the RCA is that it provides the capability of multiplex analysis so that large numbers of sequences can be analyzed in parallel. By combining the sensitivity of RCA and parallel detection on arrays, many sequences can be analyzed directly from genomic DNA.

CHEMICAL LIGATION TECHNIQUES

A variation of LCR utilizes a "chemical ligation" of sorts, as is generally outlined in U.S. Patent Nos. 5,616,464 and 5,767,259, both of which are hereby expressly incorporated by reference in their entirety. In this embodiment, similar to enzymatic ligation, a pair of primers are utilized, wherein the

first primer is substantially complementary to a first domain of the target and the second primer is substantially complementary to an adjacent second domain of the target (although, as for enzymatic ligation, if a "gap" exists, a polymerase and dNTPs may be added to "fill in" the gap). Each primer has a portion that acts as a "side chain" that does not bind the target sequence and acts as one half of a stem structure that interacts non-covalently through hydrogen bonding, salt bridges, van der Waal's forces, etc. Preferred embodiments utilize substantially complementary nucleic acids as the side chains. Thus, upon hybridization of the primers to the target sequence, the side chains of the primers are brought into spatial proximity, and, if the side chains comprise nucleic acids as well, can also form side chain hybridization complexes.

At least one of the side chains of the primers comprises an activatable cross-linking agent, generally covalently attached to the side chain, that upon activation, results in a chemical cross-link or chemical ligation. The activatible group may comprise any moiety that will allow cross-linking of the side chains, and include groups activated chemically, photonically and thermally, with photoactivatable groups being preferred. In some embodiments a single activatable group on one of the side chains is enough to result in cross-linking via interaction to a functional group on the other side chain; in alternate embodiments, activatable groups are required on each side chain.

Once the hybridization complex is formed, and the cross-linking agent has been activated such that the primers have been covalently attached, the reaction is subjected to conditions to allow for the disassocation of the hybridization complex, thus freeing up the target to serve as a template for the next ligation or cross-linking. In this way, signal amplification occurs, and can be detected as outlined herein.

INVASIVE CLEAVAGE TECHNIQUES

In a preferred embodiment, the signal amplification technique is invasive cleavage technology, which is described in a number of patents and patent applications, including U.S. Patent Nos. 5,846,717; 5,614,402; 5,719,028; 5,541,311; and 5,843,669, all of which are hereby incorporated by reference in their entirety.

Generally, invasive cleavage technology may be described as follows. A target nucleic acid is recognized by two distinct probes. A first probe, generally referred to herein as an "invader" probe, is substantially complementary to a first portion of the target nucleic acid. A second probe, generally referred to herein as a "signal probe", is partially complementary to the target nucleic acid; the 3' end of the signal oligonucleotide is substantially complementary to the target sequence while the 5' end is non-complementary and preferably forms a single-stranded "tail" or "arm". The non-complementary end of the second probe preferably comprises a "generic" or "unique" sequence, frequently referred to herein as a "detection sequence", that is used to indicate the presence or absence of the target nucleic acid, as described below. The detection sequence of the second probe preferably comprises

at least one detectable label, although as outlined herein, since this detection sequence can function as a target sequence for a capture probe, sandwich configurations utilizing label probes as described herein may also be done.

Hybridization of the first and second oligonucleotides near or adjacent to one another on the target nucleic acid forms a number of structures. In a preferred embodiment, a forked cleavage structure forms and is a substrate of a nuclease which cleaves the detection sequence from the signal oligonucleotide. The site of cleavage is controlled by the distance or overlap between the 3' end of the invader oligonucleotide and the downstream fork of the signal oligonucleotide. Therefore, neither oligonucleotide is subject to cleavage when misaligned or when unattached to target nucleic acid.

In a preferred embodiment, the nuclease that recognizes the forked cleavage structure and catalyzes release of the tail is thermostable, thereby, allowing thermal cycling of the cleavage reaction, if desired. Preferred nucleases derived from thermostable DNA polymerases that have been modified to have reduced synthetic activity which is an undesirable side-reaction during cleavage are disclosed in U.S. Patent Nos. 5,719,028 and 5,843,669, hereby expressly by reference. The synthetic activity of the DNA polymerase is reduced to a level where it does not interfere with detection of the cleavage reaction and detection of the freed tail. Preferably the DNA polymerase has no detectable polymerase activity. Examples of nucleases are those derived from *Thermus aquaticus*, *Thermus flavus*, or *Thermus thermophilus.*

In another embodiment, thermostable structure-specific nucleases are Flap endonucleases (FENs) selected from FEN-1 or FEN-2 like (e.g. XPG and RAD2 nucleases) from Archaebacterial species, for example, FEN-1 from *Methanococcus jannaschii*, *Pyrococcus furiosis*, *Pyrococcus woesei*, and *Archaeoglobus fulgidus*. (U.S. Patent No. 5,843,669 and Lyamichev *et al.* 1999. Nature Biotechnology 17:292-297; both of which are hereby expressly by reference).

In a preferred embodiment, the nuclease is *Afu*FEN1 or *Pfu*FEN1 nuclease. To cleave a forked structure, these nucleases require at least one overlapping nucleotide between the signal and invasive probes to recognize and cleave the 5' end of the signal probe. To effect cleavage the 3'-terminal nucleotide of the invader oligonucleotide is not required to be complementary to the target nucleic acid. In contast, mismatch of the signal probe one base upstream of the cleavage site prevents creation of the overlap and cleavage. The specificity of the nuclease reaction allows single nucleotide polymorphism (SNP) detection from, for example, genomic DNA, as outlined below (Lyamichev et al.).

In a preferred embodiment invasive cleavage technology is used. Invasive cleavage technology is based on structure-specific nucleases that cleave nucleic acids in a site-specific manner. Two probes are used: an "invader" probe and a "signalling" probe, that adjacently hybridize to a target sequence with overlap. For mismatch discrimination, the invader technology relies on complementarity at the

overlap position where cleavage occurs. The enzyme cleaves at the overlap, and releases the "tail" which may or may not be labeled. This can then be detected. The Invader™ technology is described in U.S. Patent Nos. 5,846,717; 5,614,402; 5,719,028; 5,541,311; and 5,843,669, all of which are hereby incorporated by reference.

The invasive cleavage assay is preferably performed on an array format. In a preferred embodiment, the signal probe has a detectable label, attached 5' from the site of nuclease cleavage (e.g. within the detection sequence) and a capture tag, as described below (e.g. biotin or other hapten) 3' from the site of nuclease cleavage. After the assay is carried out, the 3' portion of the cleaved signal probe (e.g. the the detection sequence) are extracted, for example, by binding to streptavidin beads or by crosslinking through the capture tag to produce aggregates or by antibody to an attached hapten. By "capture tag" herein is a meant one of a pair of binding partners as described above, such as antigen/antibody pairs, digoxygenenin, dinitrophenol, etc.

The cleaved 5' region, e.g. the detection sequence, of the signal probe, comprises a label and is detected and optionally quantitated. In one embodiment, the cleaved 5' region is hybridized to a probe on an array (capture probe) and optically detected. As described below, many signal probes can be analyzed in parallel by hybridization to their complementary probes in an array.

In a preferred embodiment, the invasive cleavage reaction is configured to utilize a fluorophore-quencher reaction. A signalling probe comprising both a fluorophore and a quencher is used, with the fluorophore and the quencher on opposite sides of the cleavage site. As will be appreciated by those in the art, these will be positioned closely together. Thus, in the absence of cleavage, very little signal is seen due to the quenching reaction. After cleavage, however, the distance between the two is large, and thus fluorescence can be detected. Upon assembly of an assay complex, comprising the target sequence, an invader probe, and a signalling probe, and the introduction of the cleavage enzyme, the cleavage of the complex results in the disassociation of the quencher from the complex, resulting in an increase in fluorescence.

In this embodiment, suitable fluorophore-quencher pairs are as known in the art. For example, suitable quencher molecules comprise Dabcyl.

As will be appreciated by those in the art, this system can be configured in a variety of conformations, as discussed in Figure 4.

In a preferred embodiment, to obtain higher specificity and reduce the detection of contaminating uncleaved signal probe or incorrectly cleaved product, an additional enzymatic recognition step is introduced in the array capture procedure. For example, the cleaved signal probe binds to a capture probe to produce a double-stranded nucleic acid in the array. In this embodiment, the 3' end of the

cleaved signal probe is adjacent to the 5' end of one strand of the capture probe, thereby, forming a substrate for DNA ligase (Broude et al. 1991. PNAS 91: 3072-3076). Only correctly cleaved product is ligated to the capture probe. Other incorrectly hybridized and non-cleaved signal probes are removed, for example, by heat denaturation, high stringency washes, and other methods that disrupt base pairing.

CYCLING PROBE TECHNIQUES (CPT)

In a preferred embodiment, the signal amplification technique is CPT. CPT technology is described in a number of patents and patent applications, including U.S. Patent Nos. 5,011,769, 5,403,711, 5,660,988, and 4,876,187, and PCT published applications WO 95/05480, WO 95/1416, and WO 95/00667, and U.S.S.N. 09/014,304, all of which are expressly incorporated by reference in their entirety.

Generally, CPT may be described as follows. A CPT primer (also sometimes referred to herein as a "scissile primer"), comprises two probe sequences separated by a scissile linkage. The CPT primer is substantially complementary to the target sequence and thus will hybridize to it to form a hybridization complex. The scissile linkage is cleaved, without cleaving the target sequence, resulting in the two probe sequences being separated. The two probe sequences can thus be more easily disassociated from the target, and the reaction can be repeated any number of times. The cleaved primer is then detected as outlined herein.

By "scissile linkage" herein is meant a linkage within the scissile probe that can be cleaved when the probe is part of a hybridization complex, that is, when a double-stranded complex is formed. It is important that the scissile linkage cleave only the scissile probe and not the sequence to which it is hybridized (i.e. either the target sequence or a probe sequence), such that the target sequence may be reused in the reaction for amplification of the signal. As used herein, the scissile linkage, is any connecting chemical structure which joins two probe sequences and which is capable of being selectively cleaved without cleavage of either the probe sequences or the sequence to which the scissile probe is hybridized. The scissile linkage may be a single bond, or a multiple unit sequence. As will be appreciated by those in the art, a number of possible scissile linkages may be used.

In a preferred embodiment, the scissile linkage comprises RNA. This system, previously described in as outlined above, is based on the fact that certain double-stranded nucleases, particularly ribonucleases, will nick or excise RNA nucleosides from a RNA:DNA hybridization complex. Of particular use in this embodiment is RNAseH, Exo III, and reverse transcriptase.

In one embodiment, the entire scissile probe is made of RNA, the nicking is facilitated especially when carried out with a double-stranded ribonuclease, such as RNAseH or Exo III. RNA probes made entirely of RNA sequences are particularly useful because first, they can be more easily produced

enzymatically, and second, they have more cleavage sites which are accessible to nicking or cleaving by a nicking agent, such as the ribonucleases. Thus, scissile probes made entirely of RNA do not rely on a scissile linkage since the scissile linkage is inherent in the probe.

In a preferred embodiment, when the scissile linkage is a nucleic acid such as RNA, the methods of the invention may be used to detect mismatches, as is generally described in U.S. Patent Nos. 5,660,988, and WO 95/14106, hereby expressly incorporated by reference. These mismatch detection methods are based on the fact that RNAseH may not bind to and/or cleave an RNA:DNA duplex if there are mismatches present in the sequence. Thus, in the NA_1-R-NA_2 embodiments, NA_1 and NA_2 are non-RNA nucleic acids, preferably DNA. Preferably, the mismatch is within the RNA:DNA duplex, but in some embodiments the mismatch is present in an adjacent sequence very close to the desired sequence, close enough to affect the RNAseH (generally within one or two bases). Thus, in this embodiment, the nucleic acid scissile linkage is designed such that the sequence of the scissile linkage reflects the particular sequence to be detected, i.e. the area of the putative mismatch.

In some embodiments of mismatch detection, the rate of generation of the released fragments is such that the methods provide, essentially, a yes/no result, whereby the detection of virtually any released fragment indicates the presence of the desired target sequence. Typically, however, when there is only a minimal mismatch (for example, a 1-, 2- or 3-base mismatch, or a 3-base deletion), there is some generation of cleaved sequences even though the target sequence is not present. Thus, the rate of generation of cleaved fragments, and/or the final amount of cleaved fragments, is quantified to indicate the presence or absence of the target. In addition, the use of secondary and tertiary scissile probes may be particularly useful in this embodiment, as this can amplify the differences between a perfect match and a mismatch. These methods may be particularly useful in the determination of homozygotic or heterozygotic states of a patient.

In this embodiment, it is an important feature of the scissile linkage that its length is determined by the suspected difference between the target and the probe. In particular, this means that the scissile linkage must be of sufficient length to encompass the suspected difference, yet short enough so that the scissile linkage cannot inappropriately "specifically hybridize" to the selected nucleic acid molecule when the suspected difference is present; such inappropriate hybridization would permit excision and thus cleavage of scissile linkages even though the selected nucleic acid molecule was not fully complementary to the nucleic acid probe. Thus in a preferred embodiment, the scissile linkage is between 3 to 5 nucleotides in length, such that a suspected nucleotide difference from 1 nucleotide to 3 nucleotides is encompassed by the scissile linkage, and 0, 1 or 2 nucleotides are on either side of the difference.

Thus, when the scissile linkage is nucleic acid, preferred embodiments utilize from 1 to about 100 nucleotides, with from about 2 to about 20 being preferred and from about 5 to about 10 being

particularly preferred.

CPT may be done enzymatically or chemically. That is, in addition to RNAseH, there are several other cleaving agents which may be useful in cleaving RNA (or other nucleic acid) scissile bonds. For example, several chemical nucleases have been reported; see for example Sigman et al., Annu. Rev. Biochem. 1990, 59, 207-236; Sigman et al., Chem. Rev. 1993, 93, 2295-2316; Bashkin et al., J. Org. Chem. 1990, 55, 5125-5132; and Sigman et al., Nucleic Acids and Molecular Biology, vol. 3, F. Eckstein and D.M.J. Lilley (Eds), Springer-Verlag, Heidelberg 1989, pp. 13-27; all of which are hereby expressly incorporated by reference.

Specific RNA hydrolysis is also an active area; see for example Chin, Acc. Chem. Res. 1991, 24, 145-152; Breslow et al., Tetrahedron, 1991, 47, 2365-2376; Anslyn et al., Angew. Chem. Int. Ed. Engl., 1997, 36, 432-450; and references therein, all of which are expressly incorporated by reference. Reactive phosphate centers are also of interest in developing scissile linkages, see Hendry et al., Prog. Inorg. Chem. : Bioinorganic Chem. 1990, 31, 201-258 also expressly incorporated by reference.

Current approaches to site-directed RNA hydrolysis include the conjugation of a reactive moiety capable of cleaving phosphodiester bonds to a recognition element capable of sequence-specifically hybridizing to RNA. In most cases, a metal complex is covalently attached to a DNA strand which forms a stable heteroduplex. Upon hybridization, a Lewis acid is placed in close proximity to the RNA backbone to effect hydrolysis; see Magda et al., J. Am. Chem. Soc. 1994, 116, 7439; Hall et al., Chem. Biology 1994, 1, 185-190; Bashkin et al., J. Am. Chem. Soc. 1994, 116, 5981-5982; Hall et al., Nucleic Acids Res. 1996, 24, 3522; Magda et al., J. Am. Chem. Soc. 1997, 119, 2293; and Magda et al., J. Am. Chem. Soc. 1997, 119, 6947, all of which are expressly incorporated by reference.

In a similar fashion, DNA-polyamine conjugates have been demonstrated to induce site-directed RNA strand scission; see for example, Yoshinari et al., J. Am. Chem. Soc. 1991, 113, 5899-5901; Endo et al., J. Org. Chem. 1997, 62, 846; and Barbier et al., J. Am. Chem. Soc. 1992, 114, 3511-3515, all of which are expressly incorporated by reference.

In a preferred embodiment, the scissile linkage is not necessarily RNA. For example, chemical cleavage moieties may be used to cleave basic sites in nucleic acids; see Belmont, et al.,New J. Chem. 1997, 21, 47-54; and references therein, all of which are expressly incorporated herein by reference. Similarly, photocleavable moieties, for example, using transition metals, may be used; see Moucheron, et al., Inorg. Chem. 1997, 36, 584-592, hereby expressly by reference.

Other approaches rely on chemical moieties or enzymes; see for example Keck et al., Biochemistry 1995, 34, 12029-12037; Kirk et al., Chem. Commun. 1998, in press; cleavage of G-U basepairs by metal complexes; see Biochemistry, 1992, 31, 5423-5429; diamine complexes for cleavage of RNA;

Komiyama, et al., J. Org. Chem. 1997, 62, 2155-2160; and Chow et al., Chem. Rev. 1997, 97, 1489-1513, and references therein, all of which are expressly incorporated herein by reference.

The first step of the CPT method requires hybridizing a primary scissile primer (also called a primary scissile probe) to the target. This is preferably done at a temperature that allows both the binding of the longer primary probe and disassociation of the shorter cleaved portions of the primary probe, as will be appreciated by those in the art. As outlined herein, this may be done in solution, or either the target or one or more of the scissile probes may be attached to a solid support. For example, it is possible to utilize "anchor probes" on a solid support which are substantially complementary to a portion of the target sequence, preferably a sequence that is not the same sequence to which a scissile probe will bind.

Similarly, as outlined herein, a preferred embodiment has one or more of the scissile probes attached to a solid support such as a bead. In this embodiment, the soluble target diffuses to allow the formation of the hybridization complex between the soluble target sequence and the support-bound scissile probe. In this embodiment, it may be desirable to include additional scissile linkages in the scissile probes to allow the release of two or more probe sequences, such that more than one probe sequence per scissile probe may be detected, as is outlined below, in the interests of maximizing the signal.

In this embodiment (and in other amplification techniques herein), preferred methods utilize cutting or shearing techniques to cut the nucleic acid sample containing the target sequence into a size that will allow sufficient diffusion of the target sequence to the surface of a bead. This may be accomplished by shearing the nucleic acid through mechanical forces (e.g. sonication) or by cleaving the nucleic acid using restriction endonucleases. Alternatively, a fragment containing the target may be generated using polymerase, primers and the sample as a template, as in polymerase chain reaction (PCR). In addition, amplification of the target using PCR or LCR or related methods may also be done; this may be particularly useful when the target sequence is present in the sample at extremely low copy numbers. Similarly, numerous techniques are known in the art to increase the rate of mixing and hybridization including agitation, heating, techniques that increase the overall concentration such as precipitation, drying, dialysis, centrifugation, electrophoresis, magnetic bead concentration, etc.

In general, the scissile probes are introduced in a molar excess to their targets (including both the target sequence or other scissile probes, for example when secondary or tertiary scissile probes are used), with ratios of scissile probe:target of at least about 100:1 being preferred, at least about 1000:1 being particularly preferred, and at least about 10,000:1 being especially preferred. In some embodiments the excess of probe:target will be much greater. In addition, ratios such as these may be used for all the amplification techniques outlined herein.

Once the hybridization complex between the primary scissile probe and the target has been formed, the complex is subjected to cleavage conditions. As will be appreciated, this depends on the composition of the scissile probe; if it is RNA, RNAseH is introduced. It should be noted that under certain circumstances, such as is generally outlined in WO 95/00666 and WO 95/00667, hereby incorporated by reference, the use of a double-stranded binding agent such as RNAseH may allow the reaction to proceed even at temperatures above the Tm of the primary probe:target hybridization complex. Accordingly, the addition of scissile probe to the target can be done either first, and then the cleavage agent or cleavage conditions introduced, or the probes may be added in the presence of the cleavage agent or conditions.

The cleavage conditions result in the separation of the two (or more) probe sequences of the primary scissile probe. As a result, the shorter probe sequences will no longer remain hybridized to the target sequence, and thus the hybridization complex will disassociate, leaving the target sequence intact.

The optimal temperature for carrying out the CPT reactions is generally from about 5°C to about 25°C below the melting temperatures of the probe:target hybridization complex. This provides for a rapid rate of hybridization and high degree of specificity for the target sequence. The Tm of any particular hybridization complex depends on salt concentration, G-C content, and length of the complex, as is known in the art and described herein.

During the reaction, as for the other amplification techniques herein, it may be necessary to suppress cleavage of the probe, as well as the target sequence, by nonspecific nucleases. Such nucleases are generally removed from the sample during the isolation of the DNA by heating or extraction procedures. A number of inhibitors of single-stranded nucleases such as vanadate, inhibitors it-ACE and RNAsin, a placental protein, do not affect the activity of RNAseH. This may not be necessary depending on the purity of the RNAseH and/or the target sample.

These steps are repeated by allowing the reaction to proceed for a period of time. The reaction is usually carried out for about 15 minutes to about 1 hour. Generally, each molecule of the target sequence will turnover between 100 and 1000 times in this period, depending on the length and sequence of the probe, the specific reaction conditions, and the cleavage method. For example, for each copy of the target sequence present in the test sample 100 to 1000 molecules will be cleaved by RNAseH. Higher levels of amplification can be obtained by allowing the reaction to proceed longer, or using secondary, tertiary, or quaternary probes, as is outlined herein.

Upon completion of the reaction, generally determined by time or amount of cleavage, the uncleaved scissile probes must be removed or neutralized prior to detection, such that the uncleaved probe does not bind to a detection probe, causing false positive signals. This may be done in a variety of ways, as is generally described below.

In a preferred embodiment, the separation is facilitated by the use of beads containing the primary probe. Thus, when the scissile probes are attached to beads, removal of the beads by filtration, centrifugation, the application of a magnetic field, electrostatic interactions for charged beads, adhesion, etc., results in the removal of the uncleaved probes.

In a preferred embodiment, the separation is based on strong acid precipitation. This is useful to separate long (generally greater than 50 nucleotides) from smaller fragments (generally about 10 nucleotides). The introduction of a strong acid such as trichloroacetic acid into the solution causes the longer probe to precipitate, while the smaller cleaved fragments remain in solution. The solution can be centrifuged or filtered to remove the precipitate, and the cleaved probe sequences can be quantitated.

In a preferred embodiment, the scissile probe contains both a detectable label and an affinity binding ligand or moiety, such that an affinity support is used to carry out the separation. In this embodiment, it is important that the detectable label used for detection is not on the same probe sequence that contains the affinity moiety, such that removal of the uncleaved probe, and the cleaved probe containing the affinity moiety, does not remove all the detectable labels. Alternatively, the scissile probe may contain a capture tag; the binding partner of the capture tag is attached to a solid support such as glass beads, latex beads, dextrans, etc. and used to pull out the uncleaved probes, as is known in the art. The cleaved probe sequences, which do not contain the capture tag, remain in solution and then can be detected as outlined below.

In a preferred embodiment, similar to the above embodiment, a separation sequence of nucleic acid is included in the scissile probe, which is not cleaved during the reaction. A nucleic acid complementary to the separation sequence is attached to a solid support such as a bead and serves as a catcher sequence. Preferably, the separation sequence is added to the scissile probes, and is not recognized by the target sequence, such that a generalized catcher sequence may be utilized in a variety of assays.

After removal of the uncleaved probe, as required, detection proceeds via the addition of the cleaved probe sequences to the array compositions, as outlined below. In general, the cleaved probe is bound to a capture probe, either directly or indirectly, and the label is detected. In a preferred embodiment, no higher order probes are used, and detection is based on the probe sequence(s) of the primary primer. In a preferred embodiment, at least one, and preferably more, secondary probes (also referred to herein as secondary primers) are used; the secondary probes hybridize to the domains of the cleavage probes; etc.

Thus, CPT requires, again in no particular order, a first CPT primer comprising a first probe sequence, a scissile linkage and a second probe sequence; and a cleavage agent.

In this manner, CPT results in the generation of a large amount of cleaved primers, which then can be detected as outlined below.

Labeling techniques

In general, either direct or indirect detection of the target products can be done. "Direct" detection as used in this context, as for the other amplification strategies outlined herein, requires the incorporation of a label, in this case a detectable label, preferably an optical label such as a fluorophore, into the target sequence, with detection proceeding as outlined below. In this embodiment, the label(s) may be incorporated in three ways: (1) the primers comprise the label(s), for example attached to the base, a ribose, a phosphate, or to analogous structures in a nucleic acid analog; (2) modified nucleosides are used that are modified at either the base or the ribose (or to analogous structures in a nucleic acid analog) with the label(s); these label-modified nucleosides are then converted to the triphosphate form and are incorporated into the newly synthesized strand by a polymerase; (3) modified nucleotides are used that comprise a functional group that can be used to add a detectable label; or (4) modified primers are used that comprise a functional group that can be used to add a detectable label. Any of these methods result in a newly synthesized strand that comprises labels, that can be directly detected as outlined below.

Thus, the modified strands comprise a detection label. By "detection label" or "detectable label" herein is meant a moiety that allows detection. This may be a primary label or a secondary label.

In a preferred embodiment, the detection label is a primary label. A primary label is one that can be directly detected, such as a fluorophore. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) magnetic, electrical, thermal labels; and c) colored or luminescent dyes. Labels can also include enzymes (horseradish peroxidase, etc.) and magnetic particles. Preferred labels include chromophores or phosphors but are preferably fluorescent dyes. Suitable dyes for use in the invention include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, alexa dyes, phycoerythin, bodipy, and others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference.

In a preferred embodiment, a secondary detectable label is used. Accordingly, detection labels may be primary labels (i.e. directly detectable) or secondary labels (indirectly detectable). A secondary label is one that is indirectly detected; for example, a secondary label can bind or react with a primary label for detection, or may allow the separation of the compound comprising the secondary label from unlabeled materials, etc. Secondary labels find particular use in systems requiring separation of labeled and unlabeled probes, such as SBE reactions. Secondary labels include, but are not limited to, one of a binding partner pair; chemically modifiable moieties; nuclease inhibitors, etc.

In a preferred embodiment, the secondary label is a binding partner pair. For example, the label may be a hapten or antigen, which will bind its binding partner. In a preferred embodiment, the binding partner can be attached to a solid support to allow separation of extended and non-extended primers. For example, suitable binding partner pairs include, but are not limited to: antigens (such as proteins (including peptides)) and antibodies (including fragments thereof (FAbs, etc.)); proteins and small molecules, including biotin/streptavidin; enzymes and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partners. Nucleic acid - nucleic acid binding proteins pairs are also useful. In general, the smaller of the pair is attached to the NTP for incorporation into the extension primer.

In a preferred embodiment, the binding partner pair comprises biotin or imino-biotin and streptavidin. Imino-biotin is particularly preferred as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95°C).

In a preferred embodiment, the binding partner pair comprises a primary detection label (for example, attached to the NTP and therefore to the extended primer) and an antibody that will specifically bind to the primary detection label. By "specifically bind" herein is meant that the partners bind with specificity sufficient to differentiate between the pair and other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the assay, including wash steps to remove non-specific binding. In some embodiments, the dissociation constants of the pair will be less than about 10^{-4}-10^{-6} M^{-1}, with less than about 10^{-5} to 10^{-9} M^{-1} being preferred and less than about 10^{-7} -10^{-9} M^{-1} being particularly preferred.

For removal of unextended primers, it is preferred that the other half of the binding pair is attached to a solid support. In this embodiment, the solid support may be any as described herein for substrates and microspheres, and the form is preferably microspheres as well; for example, a preferred embodiment utilizes magnetic beads that can be easily introduced to the sample and easily removed, although any affinity chromatography formats may be used as well. Standard methods are used to attach the binding partner to the solid support, and can include direct or indirect attachment methods. For example, biotin labeled antibodies to fluorophores can be attached to streptavidin coated magnetic beads.

Thus, in this embodiment, the extended primers comprise a binding partner that is contacted with its binding partner under conditions wherein the extended primers are separated from the unextended primers. These extended primers can then be added to the array comprising capture probes as described herein.

In a preferred embodiment, the secondary label is a chemically modifiable moiety. In this embodiment, labels comprising reactive functional groups are incorporated into the nucleic acid. The functional

group can then be subsequently labeled with a primary label. Suitable functional groups include, but are not limited to, amino groups, carboxy groups, maleimide groups, oxo groups and thiol groups, with amino groups and thiol groups being particularly preferred.

Removal of unextended primers

In a preferred embodiment, it is desirable to remove the unextended or unreacted primers from the assay mixture, and particularly from the array, as unextended primers will compete with the extended (labeled) primers in binding to capture probes, thereby diminishing the signal. The concentration of the unextended primers relative to the extended primer may be relatively high, since a large excess of primer is usually required to generate efficient primer annealing. Accordingly, a number of different techniques may be used to facilitate the removal of unextended primers. While the discussion below applies specifically to SBE, these techniques may be used in any of the methods described herein.

In a preferred embodiment, the NTPs (or, in the case of other methods, one or more of the probes) comprise a secondary detectable label that can be used to separate extended and non-extended primers. As outlined above, detection labels may be primary labels (i.e. directly detectable) or secondary labels (indirectly detectable). A secondary label is one that is indirectly detected; for example, a secondary label can bind or react with a primary label for detection, or may allow the separation of the compound comprising the secondary label from unlabeled materials, etc. Secondary labels find particular use in systems requiring separation of labeled and unlabeled probes, such as SBE, OLA, invasive cleavage, etc. reactions; in addition, these techniques may be used with many of the other techniques described herein. Secondary labels include, but are not limited to, one of a binding partner pair; chemically modifiable moieties; nuclease inhibitors, etc.

In a preferred embodiment, the secondary label is a binding partner pair. For example, the label may be a hapten or antigen, which will bind its binding partner (generally attached to a solid support) and thus allow separation of extended and non-extended primers. For example, suitable binding partner pairs include, but are not limited to: antigens (such as proteins (including peptides)) and antibodies (including fragments thereof (FAbs, etc.)); proteins and small molecules, including biotin/streptavidin and digoxygenin and antibodies; enzymes and substrates or inhibitors; other protein-protein interacting pairs; receptor-ligands; and carbohydrates and their binding partners, are also suitable binding pairs. Nucleic acid - nucleic acid binding proteins pairs are also useful. In general, the smaller of the pair is attached to the NTP (or the probe) for incorporation into the extension primer.

In a preferred embodiment, the binding partner pair comprises biotin or imino-biotin and streptavidin. Imino-biotin is particularly preferred when the methods require the later separation of the pair, as imino-biotin disassociates from streptavidin in pH 4.0 buffer while biotin requires harsh denaturants (e.g. 6 M guanidinium HCl, pH 1.5 or 90% formamide at 95°C).

In addition, the use of streptavidin/biotin systems can be used to separate unreacted and reacted probes (for example in SBE, invasive cleavage, etc.). For example, the addition of streptavidin to a nucleic acid greatly increases its size, as well as changes its physical properties, to allow more efficient separation techniques. For example, the mixtures can be size fractionated by exclusion chromatography, affinity chromatography, filtration or differential precipitation. Alternatively, an 3' exonuclease may be added to a mixture of 3' labeled biotin/streptavidin; only the unreacted oligonucleotides will be degraded. Following exonuclease treatment, the exonuclease and the streptavidin can be degraded using a protease such as proteinase K. The surviving nucleic acids (i.e. those that were biotinylated) are then hybridized to the array.

In a preferred embodiment, the binding partner pair comprises a primary detection label (attached to the NTP and therefore to the extended primer) and an antibody that will specifically bind to the primary detection label. By "specifically bind" herein is meant that the partners bind with specificity sufficient to differentiate between the pair and other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the assay, including wash steps to remove non-specific binding. In some embodiments, the dissociation constants of the pair will be less than about 10^{-4}-10^{-6} M^{-1}, with less than about 10^{-5} to 10^{-9} M^{-1} being preferred and less than about 10^{-7} - 10^{-9} M^{-1} being particularly preferred.

In this embodiment, it is preferred that the other half of the binding pair is attached to a solid support. In this embodiment, the solid support may be any as described herein for substrates and microspheres, and the form is preferably microspheres as well; for example, a preferred embodiment utilizes magnetic beads that can be easily introduced to the sample and easily removed, although any affinity chromatography formats may be used as well. Standard methods are used to attach the binding partner to the solid support, and can include direct or indirect attachment methods. For example, biotin labeled antibodies to fluorophores can be attached to streptavidin coated magnetic beads.

Thus, in this embodiment, the extended primers comprise a binding member that is contacted with its binding partner under conditions wherein the extended primers are separated from the unextended primers. These extended primers can then be added to the array comprising capture probes as described herein.

In a preferred embodiment, the secondary label is a chemically modifiable moiety. In this embodiment, labels comprising reactive functional groups are incorporated into the nucleic acid.

In a preferred embodiment, the secondary label is a nuclease inhibitor. In this embodiment, the chain-terminating NTPs are chosen to render extended primers resistant to nucleases, such as 3'-exonucleases. Addition of an exonuclease will digest the non-extended primers leaving only the

extended primers to bind to the capture probes on the array. This may also be done with OLA, wherein the ligated probe will be protected but the unprotected ligation probe will be digested.

In this embodiment, suitable 3'-exonucleases include, but are not limited to, exo I, exo III, exo VII, etc.

SANDWICH ASSAY TECHNIQUES

In a preferred embodiment, the signal amplification technique is a "sandwich" assay, as is generally described in U.S.S.N. 60/073,011 and in U.S. Patent Nos. 5,681,702, 5,597,909, 5,545,730, 5,594,117, 5,591,584, 5,571,670, 5,580,731, 5,571,670, 5,591,584, 5,624,802, 5,635,352, 5,594,118, 5,359,100, 5,124,246 and 5,681,697, all of which are hereby incorporated by reference. Although sandwich assays do not result in the alteration of primers, sandwich assays can be considered signal amplification techniques since multiple signals (i.e. label probes) are bound to a single target, resulting in the amplification of the signal. Sandwich assays may be used when the target sequence does not contain a label; or when adapters are used, as outlined below.

As discussed herein, it should be noted that the sandwich assays can be used for the detection of primary target sequences (e.g. from a patient sample), or as a method to detect the product of an amplification reaction as outlined above; thus for example, any of the newly synthesized strands outlined above, for example using PCR, LCR, NASBA, SDA, etc., may be used as the "target sequence" in a sandwich assay.

As will be appreciated by those in the art, the systems of the invention may take on a large number of different configurations. In general, there are three types of systems that can be used: (1) "non-sandwich" systems (also referred to herein as "direct" detection) in which the target sequence itself is labeled with detectable labels (again, either because the primers comprise labels or due to the incorporation of labels into the newly synthesized strand); (2) systems in which label probes directly bind to the target sequences; and (3) systems in which label probes are indirectly bound to the target sequences, for example through the use of amplifier probes.

The anchoring of the target sequence to the bead is done through the use of capture probes and optionally either capture extender probes (sometimes referred to as "adapter sequences" herein). When only capture probes are utilized, it is necessary to have unique capture probes for each target sequence; that is, the surface must be customized to contain unique capture probes; e.g. each bead comprises a different capture probe. Alternatively, capture extender probes may be used, that allow a "universal" surface, i.e. a surface containing a single type of capture probe that can be used to detect any target sequence. "Capture extender" probes have a first portion that will hybridize to all or part of the capture probe, and a second portion that will hybridize to a first portion of the target sequence. This then allows the generation of customized soluble probes, which as will be appreciated by those in the art is generally simpler and less costly. As shown herein, two capture extender probes may be

used. This has generally been done to stabilize assay complexes for example when the target sequence is large, or when large amplifier probes (particularly branched or dendrimer amplifier probes) are used.

Detection of the amplification reactions of the invention, including the direct detection of amplification products and indirect detection utilizing label probes (i.e. sandwich assays), is preferably done by detecting assay complexes comprising detectable labels, which can be attached to the assay complex in a variety of ways, as is more fully described below.

Once the target sequence has preferably been anchored to the array, an amplifier probe is hybridized to the target sequence, either directly, or through the use of one or more label extender probes, which serves to allow "generic" amplifier probes to be made. As for all the steps outlined herein, this may be done simultaneously with capturing, or sequentially. Preferably, the amplifier probe contains a multiplicity of amplification sequences, although in some embodiments, as described below, the amplifier probe may contain only a single amplification sequence, or at least two amplification sequences. The amplifier probe may take on a number of different forms; either a branched conformation, a dendrimer conformation, or a linear "string" of amplification sequences. Label probes comprising detectable labels (preferably but not required to be fluorophores) then hybridize to the amplification sequences (or in some cases the label probes hybridize directly to the target sequence), and the labels detected, as is more fully outlined below.

Accordingly, the present invention provides compositions comprising an amplifier probe. By "amplifier probe" or "nucleic acid multimer" or "amplification multimer" or grammatical equivalents herein is meant a nucleic acid probe that is used to facilitate signal amplification. Amplifier probes comprise at least a first single-stranded nucleic acid probe sequence, as defined below, and at least one single-stranded nucleic acid amplification sequence, with a multiplicity of amplification sequences being preferred.

Amplifier probes comprise a first probe sequence that is used, either directly or indirectly, to hybridize to the target sequence. That is, the amplifier probe itself may have a first probe sequence that is substantially complementary to the target sequence, or it has a first probe sequence that is substantially complementary to a portion of an additional probe, in this case called a label extender probe, that has a first portion that is substantially complementary to the target sequence. In a preferred embodiment, the first probe sequence of the amplifier probe is substantially complementary to the target sequence.

In general, as for all the probes herein, the first probe sequence is of a length sufficient to give specificity and stability. Thus generally, the probe sequences of the invention that are designed to hybridize to another nucleic acid (i.e. probe sequences, amplification sequences, portions or domains

of larger probes) are at least about 5 nucleosides long, with at least about 10 being preferred and at least about 15 being especially preferred.

In a preferred embodiment, several different amplifier probes are used, each with first probe sequences that will hybridize to a different portion of the target sequence. That is, there is more than one level of amplification; the amplifier probe provides an amplification of signal due to a multiplicity of labelling events, and several different amplifier probes, each with this multiplicity of labels, for each target sequence is used. Thus, preferred embodiments utilize at least two different pools of amplifier probes, each pool having a different probe sequence for hybridization to different portions of the target sequence; the only real limitation on the number of different amplifier probes will be the length of the original target sequence. In addition, it is also possible that the different amplifier probes contain different amplification sequences, although this is generally not preferred.

In a preferred embodiment, the amplifier probe does not hybridize to the sample target sequence directly, but instead hybridizes to a first portion of a label extender probe. This is particularly useful to allow the use of "generic" amplifier probes, that is, amplifier probes that can be used with a variety of different targets. This may be desirable since several of the amplifier probes require special synthesis techniques. Thus, the addition of a relatively short probe as a label extender probe is preferred. Thus, the first probe sequence of the amplifier probe is substantially complementary to a first portion or domain of a first label extender single-stranded nucleic acid probe. The label extender probe also contains a second portion or domain that is substantially complementary to a portion of the target sequence. Both of these portions are preferably at least about 10 to about 50 nucleotides in length, with a range of about 15 to about 30 being preferred. The terms "first" and "second" are not meant to confer an orientation of the sequences with respect to the 5'-3' orientation of the target or probe sequences. For example, assuming a 5'-3' orientation of the complementary target sequence, the first portion may be located either 5' to the second portion, or 3' to the second portion. For convenience herein, the order of probe sequences are generally shown from left to right.

In a preferred embodiment, more than one label extender probe-amplifier probe pair may be used, that is, n is more than 1. That is, a plurality of label extender probes may be used, each with a portion that is substantially complementary to a different portion of the target sequence; this can serve as another level of amplification. Thus, a preferred embodiment utilizes pools of at least two label extender probes, with the upper limit being set by the length of the target sequence.

In a preferred embodiment, more than one label extender probe is used with a single amplifier probe to reduce non-specific binding, as is generally outlined in U.S. Patent No. 5,681,697, incorporated by reference herein. In this embodiment, a first portion of the first label extender probe hybridizes to a first portion of the target sequence, and the second portion of the first label extender probe hybridizes to a first probe sequence of the amplifier probe. A first portion of the second label extender probe

hybridizes to a second portion of the target sequence, and the second portion of the second label extender probe hybridizes to a second probe sequence of the amplifier probe. These form structures sometimes referred to as "cruciform" structures or configurations, and are generally done to confer stability when large branched or dendrimeric amplifier probes are used.

In addition, as will be appreciated by those in the art, the label extender probes may interact with a preamplifier probe, described below, rather than the amplifier probe directly.

Similarly, as outlined above, a preferred embodiment utilizes several different amplifier probes, each with first probe sequences that will hybridize to a different portion of the label extender probe. In addition, as outlined above, it is also possible that the different amplifier probes contain different amplification sequences, although this is generally not preferred.

In addition to the first probe sequence, the amplifier probe also comprises at least one amplification sequence. An "amplification sequence" or "amplification segment" or grammatical equivalents herein is meant a sequence that is used, either directly or indirectly, to bind to a first portion of a label probe as is more fully described below. Preferably, the amplifier probe comprises a multiplicity of amplification sequences, with from about 3 to about 1000 being preferred, from about 10 to about 100 being particularly preferred, and about 50 being especially preferred. In some cases, for example when linear amplifier probes are used, from 1 to about 20 is preferred with from about 5 to about 10 being particularly preferred.

The amplification sequences may be linked to each other in a variety of ways, as will be appreciated by those in the art. They may be covalently linked directly to each other, or to intervening sequences or chemical moieties, through nucleic acid linkages such as phosphodiester bonds, PNA bonds, etc., or through interposed linking agents such amino acid, carbohydrate or polyol bridges, or through other cross-linking agents or binding partners. The site(s) of linkage may be at the ends of a segment, and/or at one or more internal nucleotides in the strand. In a preferred embodiment, the amplification sequences are attached via nucleic acid linkages.

In a preferred embodiment, branched amplifier probes are used, as are generally described in U.S. Patent No. 5,124,246, hereby incorporated by reference. Branched amplifier probes may take on "fork-like" or "comb-like" conformations. "Fork-like" branched amplifier probes generally have three or more oligonucleotide segments emanating from a point of origin to form a branched structure. The point of origin may be another nucleotide segment or a multifunctional molecule to whcih at least three segments can be covalently or tightly bound. "Comb-like" branched amplifier probes have a linear backbone with a multiplicity of sidechain oligonucleotides extending from the backbone. In either conformation, the pendant segments will normally depend from a modified nucleotide or other organic moiety having the appropriate functional groups for attachment of oligonucleotides. Furthermore, in

either conformation, a large number of amplification sequences are available for binding, either directly or indirectly, to detection probes. In general, these structures are made as is known in the art, using modified multifunctional nucleotides, as is described in U.S. Patent Nos. 5,635,352 and 5,124,246, among others.

In a preferred embodiment, dendrimer amplifier probes are used, as are generally described in U.S. Patent No. 5,175,270, hereby expressly incorporated by reference. Dendrimeric amplifier probes have amplification sequences that are attached via hybridization, and thus have portions of double-stranded nucleic acid as a component of their structure. The outer surface of the dendrimer amplifier probe has a multiplicity of amplification sequences.

In a preferred embodiment, linear amplifier probes are used, that have individual amplification sequences linked end-to-end either directly or with short intervening sequences to form a polymer. As with the other amplifier configurations, there may be additional sequences or moieties between the amplification sequences. In one embodiment, the linear amplifier probe has a single amplification sequence.

In addition, the amplifier probe may be totally linear, totally branched, totally dendrimeric, or any combination thereof.

The amplification sequences of the amplifier probe are used, either directly or indirectly, to bind to a label probe to allow detection. In a preferred embodiment, the amplification sequences of the amplifier probe are substantially complementary to a first portion of a label probe. Alternatively, amplifier extender probes are used, that have a first portion that binds to the amplification sequence and a second portion that binds to the first portion of the label probe.

In addition, the compositions of the invention may include "preamplifier" molecules, which serves a bridging moiety between the label extender molecules and the amplifier probes. In this way, more amplifier and thus more labels are ultimately bound to the detection probes. Preamplifier molecules may be either linear or branched, and typically contain in the range of about 30-3000 nucleotides.

Thus, label probes are either substantially complementary to an amplification sequence or to a portion of the target sequence.

Detection of the amplification reactions of the invention, including the direct detection of amplification products and indirect detection utilizing label probes (i.e. sandwich assays), is done by detecting assay complexes comprising labels.

ARRAYS

Detection of the amplified products described above preferably employs arrays, as defined herein. The arrays are preferably high density arrays that can allow simultaneous analysis, i.e. parallel rather than serial processing, on a number of samples. This is preferably done by forming an "array of arrays", i.e. a composite array comprising a plurality of individual arrays, that is configured to allow processing of multiple samples, as is generally outlined in U.S.S.N. 09/256,943, hereby expressly incorporated by reference. For example, each individual array is present within each well of a microtiter plate. Thus, depending on the size of the microtiter plate and the size of the individual array, very high numbers of assays can be run simultaneously; for example, using individual arrays of 2,000 and a 96 well microtiter plate, 192,000 experiments can be done at once; the same arrays in a 384 microtiter plate yields 768,000 simultaneous experiments, and a 1536 microtiter plate gives 3,072,000 experiments.

Generally, the array of array compositions of the invention can be configured in several ways. In a preferred embodiment, as is more fully outlined below, a "one component" system is used. That is, a first substrate comprising a plurality of assay locations (sometimes also referred to herein as "assay wells"), such as a microtiter plate, is configured such that each assay location contains an individual array. That is, the assay location and the array location are the same. For example, the plastic material of the microtiter plate can be formed to contain a plurality of "bead wells" in the bottom of each of the assay wells. Beads containing the capture probes of the invention can then be loaded into the bead wells in each assay location as is more fully described below.

Alternatively, a "two component" system can be used. In this embodiment, the individual arrays are formed on a second substrate, which then can be fitted or "dipped" into the first microtiter plate substrate. A preferred embodiment utilizes fiber optic bundles as the individual arrays, generally with "bead wells" etched into one surface of each individual fiber, such that the beads containing the capture probes are loaded onto the end of the fiber optic bundle. The composite array thus comprises a number of individual arrays that are configured to fit within the wells of a microtiter plate.

The present invention is generally based on previous work comprising a bead-based analytic chemistry system in which beads, also termed microspheres, carrying different chemical functionalities are distributed on a substrate comprising a patterned surface of discrete sites that can bind the individual microspheres. The beads are generally put onto the substrate randomly, and thus several different methodologies can be used to "decode" the arrays. In one embodiment, unique optical signatures are incorporated into the beads, generally fluorescent dyes, that could be used to identify the chemical functionality on any particular bead. This allows the synthesis of the nucleic acids to be divorced from their placement on an array, i.e. the capture probes may be synthesized on the beads, and then the beads are randomly distributed on a patterned surface. Since the beads are first coded with an optical signature, this means that the array can later be "decoded", i.e. after the array is made, a correlation of the location of an individual site on the array with the probe at that particular site can

be made. This means that the beads may be randomly distributed on the array, a fast and inexpensive process as compared to either the in situ synthesis or spotting techniques of the prior art. These methods are generally outlined in PCTs US98/05025 and US99/14387 and U.S.S.N.s 08/818,199 and 09/151,877, all of which are expressly incorporated herein by reference.

However, the drawback to these methods is that for a very high density array, the system requires a large number of different optical signatures, which may be difficult or time-consuming to utilize. Accordingly, the present invention also provides several improvements over these methods, generally directed to methods of coding and decoding the arrays. That is, as will be appreciated by those in the art, the placement of the probes is generally random, and thus a coding/decoding system is required to identify the probes at each location in the array. This may be done in a variety of ways, as is more fully outlined below, and generally includes: a) the use a decoding binding ligand (DBL), generally directly labeled, that binds to either the capture probes or to identifier binding ligands (IBLs) attached to the beads; b) positional decoding, for example by either targeting the placement of beads (for example by using photoactivatible or photocleavable moieties to allow the selective addition of beads to particular locations), or by using either sub-bundles or selective loading of the sites, as are more fully outlined below; c) selective decoding, wherein only those beads that bind to a target are decoded; or d) combinations of any of these. In some cases, as is more fully outlined below, this decoding may occur for all the beads, or only for those that bind a particular target analyte. Similarly, this may occur either prior to or after addition of a target analyte.

Once the identity (i.e. the actual agent) and location of each microsphere in the array has been fixed, the array is exposed to samples containing the target sequences, such as, the products of amplification reactions described above, although as outlined below, this can be done prior to or during the analysis as well. The target sequences will bind to the capture probes as is more fully outlined below, and results (in the case of optical labels) in a change in the optical signal of a particular bead.

In the present invention, "decoding" can use optical signatures, decoding binding ligands that are added during a decoding step, or a combination of these methods. The decoding binding ligands will bind either to a distinct identifier binding ligand partner that is placed on the beads, or to the capture probes, with the latter being preferred. The decoding binding ligands are either directly or indirectly labeled, and thus decoding occurs by detecting the presence of the label. By using pools of decoding binding ligands in a sequential fashion, it is possible to greatly minimize the number of required decoding steps.

Accordingly, the present invention provides composite array compositions comprising at least a first substrate with a surface comprising a plurality of assay locations. By "array" herein is meant a plurality of candidate agents in an array format; the size of the array will depend on the composition

and end use of the array. Arrays containing from about 2 different probes (i.e. different beads) to many millions can be made, with very large fiber optic arrays being possible. Generally, the array will comprise from two to as many as a billion or more per square cm, depending on the size of the beads and the substrate, as well as the end use of the array, thus very high density, high density, moderate density, low density and very low density arrays may be made. Preferred ranges for very high density arrays are from about 10,000,000 to about 2,000,000,000, (with all numbers being per square centimeter) with from about 100,000,000 to about 1,000,000,000 being preferred. High density arrays range about 100,000 to about 10,000,000, with from about 1,000,000 to about 5,000,000 being particularly preferred. Moderate density arrays range from about 10,000 to about 100,000 being particularly preferred, and from about 20,000 to about 50,000 being especially preferred. Low density arrays are generally less than 10,000, with from about 1,000 to about 5,000 being preferred. Very low density arrays are less than 1,000, with from about 10 to about 1000 being preferred, and from about 100 to about 500 being particularly preferred. In some embodiments, the compositions of the invention may not be in array format; that is, for some embodiments, compositions comprising a single bioactive agent may be made as well. In addition, in some arrays, multiple substrates may be used, either of different or identical compositions. Thus for example, large arrays may comprise a plurality of smaller substrates.

In addition, one advantage of the present compositions is that particularly through the use of fiber optic technology, extremely high density arrays can be made. Thus for example, because beads of 200 μm or less (with beads of 200 nm possible) can be used, and very small fibers are known, it is possible to have as many as 250,000 or more (in some instances, 1 million) different fibers and beads in a 1 mm^2 fiber optic bundle, with densities of greater than 15,000,000 individual beads and fibers (again, in some instances as many as 25-50 million) per 0.5 cm^2 obtainable.

By "composite array" or "combination array" or grammatical equivalents herein is meant a plurality of individual arrays, as outlined above. Generally the number of individual arrays is set by the size of the microtiter plate used; thus, 96 well, 384 well and 1536 well microtiter plates utilize composite arrays comprising 96, 384 and 1536 individual arrays, although as will be appreciated by those in the art, not each microtiter well need contain an individual array. It should be noted that the composite arrays can comprise individual arrays that are identical, similar or different. That is, in some embodiments, it may be desirable to do the same 2,000 assays on 96 different samples; alternatively, doing 192,000 experiments on the same sample (i.e. the same sample in each of the 96 wells) may be desirable. Alternatively, each row or column of the composite array could be the same, for redundancy/quality control. As will be appreciated by those in the art, there are a variety of ways to configure the system. In addition, the random nature of the arrays may mean that the same population of beads may be added to two different surfaces, resulting in substantially similar but perhaps not identical arrays.

By "substrate" or "solid support" or other grammatical equivalents herein is meant any material that

can be modified to contain discrete individual sites appropriate for the attachment or association of beads and is amenable to at least one detection method. As will be appreciated by those in the art, the number of possible substrates is very large. Possible substrates include, but are not limited to, glass and modified or functionalized glass, plastics (including acrylics, polystyrene and copolymers of styrene and other materials, polypropylene, polyethylene, polybutylene, polyurethanes, TeflonJ, etc.), polysaccharides, nylon or nitrocellulose, resins, silica or silica-based materials including silicon and modified silicon, carbon, metals, inorganic glasses, plastics, optical fiber bundles, and a variety of other polymers. In general, the substrates allow optical detection and do not themselves appreciably fluorescese.

Generally the substrate is flat (planar), although as will be appreciated by those in the art, other configurations of substrates may be used as well; for example, three dimensional configurations can be used, for example by embedding the beads in a porous block of plastic that allows sample access to the beads and using a confocal microscope for detection. Similarly, the beads may be placed on the inside surface of a tube, for flow-through sample analysis to minimize sample volume. Preferred substrates include optical fiber bundles as discussed below, and flat planar substrates such as glass, polystyrene and other plastics and acrylics.

The first substrate comprises a surface comprising a plurality of assay locations, i.e. the location where the assay for the detection of a target analyte will occur. The assay locations are generally physically separated from each other, for example as assay wells in a microtiter plate, although other configurations (hydrophobicity/hydrophilicity, etc.) can be used to separate the assay locations.

In a preferred embodiment, the second substrate is an optical fiber bundle or array, as is generally described in U.S.S.N.s 08/944,850 and 08/519,062, PCT US98/05025, and PCT US98/09163, all of which are expressly incorporated herein by reference. Preferred embodiments utilize preformed unitary fiber optic arrays. By "preformed unitary fiber optic array" herein is meant an array of discrete individual fiber optic strands that are co-axially disposed and joined along their lengths. The fiber strands are generally individually clad. However, one thing that distinguished a preformed unitary array from other fiber optic formats is that the fibers are not individually physically manipulatable; that is, one strand generally cannot be physically separated at any point along its length from another fiber strand.

In a preferred embodiment, the array comprises a plurality of discrete sites. Thus, in the former case, the assay location is the same as the array location, as described herein. In the latter case, the array location is fitted into the assay location separately. In these embodiments, at least one surface of the substrate is modified to contain discrete, individual sites for later association of microspheres. These sites may comprise physically altered sites, i.e. physical configurations such as wells or small depressions in the substrate that can retain the beads, such that a microsphere can rest in the well, or

the use of other forces (magnetic or compressive), or chemically altered or active sites, such as chemically functionalized sites, electrostatically altered sites, hydrophobically/ hydrophilically functionalized sites, spots of adhesive, etc.

The sites may be a pattern, i.e. a regular design or configuration, or randomly distributed. A preferred embodiment utilizes a regular pattern of sites such that the sites may be addressed in the X-Y coordinate plane. "Pattern" in this sense includes a repeating unit cell, preferably one that allows a high density of beads on the substrate. However, it should be noted that these sites may not be discrete sites. That is, it is possible to use a uniform surface of adhesive or chemical functionalities, for example, that allows the attachment of beads at any position. That is, the surface of the substrate is modified to allow attachment of the microspheres at individual sites, whether or not those sites are contiguous or non-contiguous with other sites. Thus, the surface of the substrate may be modified such that discrete sites are formed that can only have a single associated bead, or alternatively, the surface of the substrate is modified and beads may go down anywhere, but they end up at discrete sites.

In a preferred embodiment, the surface of the substrate is modified to contain wells, i.e. depressions in the surface of the substrate. This may be done as is generally known in the art using a variety of techniques, including, but not limited to, photolithography, stamping techniques, molding techniques and microetching techniques. As will be appreciated by those in the art, the technique used will depend on the composition and shape of the substrate. When the first substrate comprises both the assay locations and the individual arrays, a preferred method utilizes molding techniques that form the bead wells in the bottom of the assay wells in a microtiter plate.

In a preferred embodiment, physical alterations are made in a surface of the substrate to produce the sites. In a preferred embodiment, for example when the second substrate is a fiber optic bundle, the surface of the substrate is a terminal end of the fiber bundle, as is generally described in 08/818,199 and 09/151,877, both of which are hereby expressly incorporated by reference. In this embodiment, wells are made in a terminal or distal end of a fiber optic bundle comprising individual fibers. In this embodiment, the cores of the individual fibers are etched, with respect to the cladding, such that small wells or depressions are formed at one end of the fibers. The required depth of the wells will depend on the size of the beads to be added to the wells.

Generally in this embodiment, the microspheres are non-covalently associated in the wells, although the wells may additionally be chemically functionalized as is generally described below, cross-linking agents may be used, or a physical barrier may be used, i.e. a film or membrane over the beads.

In a preferred embodiment, the surface of the substrate is modified to contain chemically modified sites, that can be used to attach, either covalently or non-covalently, the microspheres of the invention

to the discrete sites or locations on the substrate. "Chemically modified sites" in this context includes, but is not limited to, the addition of a pattern of chemical functional groups including amino groups, carboxy groups, oxo groups and thiol groups, that can be used to covalently attach microspheres, which generally also contain corresponding reactive functional groups; the addition of a pattern of adhesive that can be used to bind the microspheres (either by prior chemical functionalization for the addition of the adhesive or direct addition of the adhesive); the addition of a pattern of charged groups (similar to the chemical functionalities) for the electrostatic attachment of the microspheres, i.e. when the microspheres comprise charged groups opposite to the sites; the addition of a pattern of chemical functional groups that renders the sites differentially hydrophobic or hydrophilic, such that the addition of similarly hydrophobic or hydrophilic microspheres under suitable experimental conditions will result in association of the microspheres to the sites on the basis of hydroaffinity. For example, the use of hydrophobic sites with hydrophobic beads, in an aqueous system, drives the association of the beads preferentially onto the sites. As outlined above, "pattern" in this sense includes the use of a uniform treatment of the surface to allow attachment of the beads at discrete sites, as well as treatment of the surface resulting in discrete sites. As will be appreciated by those in the art, this may be accomplished in a variety of ways.

The compositions of the invention further comprise a population of microspheres. By "population" herein is meant a plurality of beads as outlined above for arrays. Within the population are separate subpopulations, which can be a single microsphere or multiple identical microspheres. That is, in some embodiments, as is more fully outlined below, the array may contain only a single bead for each bioactive agent; preferred embodiments utilize a plurality of beads of each type.

By "microspheres" or "beads" or "particles" or grammatical equivalents herein is meant small discrete particles. The composition of the beads will vary, depending on the class of bioactive agent and the method of synthesis. Suitable bead compositions include those used in peptide, nucleic acid and organic moiety synthesis, including, but not limited to, plastics, ceramics, glass, polystyrene, methylstyrene, acrylic polymers, paramagnetic materials, thoria sol, carbon graphited, titanium dioxide, latex or cross-linked dextrans such as Sepharose, cellulose, nylon, cross-linked micelles and teflon may all be used. *"Microsphere Detection Guide"* from Bangs Laboratories, Fishers IN is a helpful guide.

The beads need not be spherical; irregular particles may be used. In addition, the beads may be porous, thus increasing the surface area of the bead available for assay. The bead sizes range from nanometers, i.e. 100 nm, to millimeters, i.e. 1 mm, with beads from about 0.2 micron to about 200 microns being preferred, and from about 0.5 to about 5 micron being particularly preferred, although in some embodiments smaller beads may be used.

It should be noted that a key component of the invention is the use of a substrate/bead pairing that

allows the association or attachment of the beads at discrete sites on the surface of the substrate, such that the beads do not move during the course of the assay.

Each microsphere comprises a capture probe although as will be appreciated by those in the art, there may be some microspheres which do not contain a capture probe, depending the on the synthetic methods. By "capture probe" or "capture nucleic acid" herein is meant a probe for the direct or indirect binding of the target sequence to a bead. By "nucleic acid" or "oligonucleotide" or grammatical equivalents herein means at least two nucleotides covalently linked together, as described above.

In a preferred embodiment, each bead comprises a single type of capture probes, although a plurality of individual probes are preferably attached to each bead. Similarly, preferred embodiments utilize more than one microsphere containing a unique capture probe; that is, there is redundancy built into the system by the use of subpopulations of microspheres, each microsphere in the subpopulation containing the same probe.

As will be appreciated by those in the art, the probes may either be synthesized directly on the beads, or they may be made and then attached after synthesis. In a preferred embodiment, linkers are used to attach the probes to the beads, to allow both good attachment, sufficient flexibility to allow good interaction with the target sequence, and to avoid undesirable binding reactions.

In a preferred embodiment, the probes are synthesized directly on the beads. As is known in the art, many classes of chemical compounds are currently synthesized on solid supports, such as peptides, organic moieties, and nucleic acids. It is a relatively straightforward matter to adjust the current synthetic techniques to use beads.

In a preferred embodiment, the probes are synthesized first, and then covalently attached to the beads. As will be appreciated by those in the art, this will be done depending on the composition of the bioactive agents and the beads. The functionalization of solid support surfaces such as certain polymers with chemically reactive groups such as thiols, amines, carboxyls, etc. is generally known in the art. Accordingly, "blank" microspheres may be used that have surface chemistries that facilitate the attachment of the desired functionality by the user. Some examples of these surface chemistries for blank microspheres include, but are not limited to, amino groups including aliphatic and aromatic amines, carboxylic acids, aldehydes, amides, chloromethyl groups, hydrazide, hydroxyl groups, sulfonates and sulfates.

In some embodiments, the beads may additionally comprise an optical signature, that can be used to identify the bioactive agent; see for example U.S.S.N.s 08/818,199 and 09/151,877, and PCT US98/05025, all of which are expressly incorporated herein by reference.

In some embodiments, the microspheres may additionally comprise identifier binding ligands for use in certain decoding systems. By "identifier binding ligands" or "IBLs" herein is meant a compound that will specifically bind a corresponding decoder binding ligand (DBL) to facilitate the elucidation of the identity of the bioactive agent attached to the bead. That is, the IBL and the corresponding DBL form a binding partner pair. By "specifically bind" herein is meant that the IBL binds its DBL with specificity sufficient to differentiate between the corresponding DBL and other DBLs (that is, DBLs for other IBLs), or other components or contaminants of the system. The binding should be sufficient to remain bound under the conditions of the decoding step, including wash steps to remove non-specific binding. In some embodiments, for example when the IBLs and corresponding DBLs are proteins or nucleic acids, the dissociation constants of the IBL to its DBL will be less than about 10^{-4}-10^{-6} M^{-1}, with less than about 10^{-5} to 10^{-9} M^{-1} being preferred and less than about 10^{-7} -10^{-9} M^{-1} being particularly preferred.

IBL-DBL binding pairs are known or can be readily found using known techniques. For example, when the IBL is a protein, the DBLs include proteins (particularly including antibodies or fragments thereof (FAbs, etc.)) or small molecules, or vice versa (the IBL is an antibody and the DBL is a protein). By "protein" herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homo-phenylalanine, citrulline and norleucine are considered amino acids for the purposes of the invention. The side chains may be in either the (R) or the (S) configuration. In the preferred embodiment, the amino acids are in the (S) or L-configuration. If non-naturally occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard *in vivo* degradations. Metal ion- metal ion ligands or chelators pairs are also useful. Antigen-antibody pairs, enzymes and substrates or inhibitors, other protein-protein interacting pairs, receptor-ligands, complementary nucleic acids, and carbohydrates and their binding partners are also suitable binding pairs. Nucleic acid - nucleic acid binding proteins pairs are also useful. Similarly, as is generally described in U.S. Patents 5,270,163, 5,475,096, 5,567,588, 5,595,877, 5,637,459, 5,683,867, 5,705,337, and related patents, hereby incorporated by reference, nucleic acid "aptomers" can be developed for binding to virtually any target; such a aptomer-target pair can be used as the IBL-DBL pair. Similarly, there is a wide body of literature relating to the development of binding pairs based on combinatorial chemistry methods.

In a preferred embodiment, the IBL is a molecule whose color or luminescence properties change in the presence of a selectively-binding DBL. For example, the IBL may be a fluorescent pH indicator whose emission intensity changes with pH. Similarly, the IBL may be a fluorescent ion indicator, whose emission properties change with ion concentration.

Alternatively, the IBL is a molecule whose color or luminescence properties change in the presence of various solvents. For example, the IBL may be a fluorescent molecule such as an ethidium salt whose fluorescence intensity increases in hydrophobic environments. Similarly, the IBL may be a derivative of fluorescein whose color changes between aqueous and nonpolar solvents.

In one embodiment, the DBL may be attached to a bead, i.e. a "decoder bead", that may carry a label such as a fluorophore.

In a preferred embodiment, the IBL-DBL pair comprise substantially complementary single-stranded nucleic acids. In this embodiment, the binding ligands can be referred to as "identifier probes" and "decoder probes". Generally, the identifier and decoder probes range from about 4 basepairs in length to about 1000, with from about 6 to about 100 being preferred, and from about 8 to about 40 being particularly preferred. What is important is that the probes are long enough to be specific, i.e. to distinguish between different IBL-DBL pairs, yet short enough to allow both a) dissociation, if necessary, under suitable experimental conditions, and b) efficient hybridization.

In a preferred embodiment, as is more fully outlined below, the IBLs do not bind to DBLs. Rather, the IBLs are used as identifier moieties ("IMs") that are identified directly, for example through the use of mass spectroscopy.

In a preferred embodiment, the microspheres may contain an optical signature. That is, as outlined in U.S.S.N.s 08/818,199 and 09/151,877, previous work had each subpopulation of microspheres comprising a unique optical signature or optical tag that is used to identify the unique capture probe of that subpopulation of microspheres; that is, decoding utilizes optical properties of the beads such that a bead comprising the unique optical signature may be distinguished from beads at other locations with different optical signatures. Thus the previous work assigned each probe a unique optical signature such that any microspheres comprising that probe are identifiable on the basis of the signature. These optical signatures comprised dyes, usually chromophores or fluorophores, that were entrapped or attached to the beads themselves. Diversity of optical signatures utilized different fluorochromes, different ratios of mixtures of fluorochromes, and different concentrations (intensities) of fluorochromes.

While generally, the present invention does not rely solely on the use of optical properties to decode the arrays, as will be appreciated by those in the art, it is possible in some embodiments to utilize optical signatures as an additional coding method, in conjunction with the present system. Thus, for example, as is more fully outlined below, the size of the array may be effectively increased while using a single set of decoding moieties in several ways, one of which is the use of optical signatures one some beads. Thus, for example, using one "set" of decoding molecules, the use of two populations of beads, one with an optical signature and one without, allows the effective doubling of the array size.

The use of multiple optical signatures similarly increases the possible size of the array.

In a preferred embodiment, each subpopulation of beads comprises a plurality of different IBLs. By using a plurality of different IBLs to encode each probe, the number of possible unique codes is substantially increased. That is, by using one unique IBL per probe, the size of the array will be the number of unique IBLs (assuming no "reuse" occurs, as outlined below). However, by using a plurality of different IBLs per bead, n, the size of the array can be increased to 2^n, when the presence or absence of each IBL is used as the indicator. For example, the assignment of 10 IBLs per bead generates a 10 bit binary code, where each bit can be designated as "1" (IBL is present) or "0" (IBL is absent). A 10 bit binary code has 2^{10} possible variants. However, as is more fully discussed below, the size of the array may be further increased if another parameter is included such as concentration or intensity; thus for example, if two different concentrations of the IBL are used, then the array size increases as 3^n. Thus, in this embodiment, each individual probe in the array is assigned a combination of IBLs, which can be added to the beads prior to the addition of the probe, after, or during the synthesis of the probe, i.e. simultaneous addition of IBLs and probes.

In some embodiment, the combination of different IBLs can be used to elucidate the sequence of the probe.

Thus, for example, using two different IBLs (IBL1 and IBL2), the first position of a nucleic acid can be elucidated: for example, adenosine can be represented by the presence of both IBL1 and IBL2; thymidine can be represented by the presence of IBL1 but not IBL2, cytosine can be represented by the presence of IBL2 but not IBL1, and guanosine can be represented by the absence of both. The second position of the nucleic acid can be done in a similar manner using IBL3 and IBL4; thus, the presence of IBL1, IBL2, IBL3 and IBL4 gives a sequence of AA; IBL1, IBL2, and IBL3 shows the sequence AT; IBL1, IBL3 and IBL4 gives the sequence TA, etc. The third position utilizes IBL5 and IBL6, etc. In this way, the use of 20 different identifiers can yield a unique code for every possible 10-mer.

In this way, a sort of "bar code" for each sequence can be constructed; the presence or absence of each distinct IBL will allow the identification of each probe.

In addition, the use of different concentrations or densities of IBLs allows a "reuse" of sorts. If, for example, the bead comprising a first agent has a 1X concentration of IBL, and a second bead comprising a second agent has a 10X concentration of IBL, using saturating concentrations of the corresponding labelled DBL allows the user to distinguish between the two beads.

Once the microspheres comprising the probes and the unique tags are generated, they are added to the substrate to form an array. In general, the methods of making the arrays and of decoding the

arrays is done to maximize the number of different candidate agents that can be uniquely encoded. The compositions of the invention may be made in a variety of ways. In general, the arrays are made by adding a solution or slurry comprising the beads to a surface containing the sites for attachment of the beads. This may be done in a variety of buffers, including aqueous and organic solvents, and mixtures. The solvent can evaporate, and excess beads removed.

In a preferred embodiment, when non-covalent methods are used to associate the beads to the array, a novel method of loading the beads onto the array is used. This method comprises exposing the array to a solution of particles (including microspheres and cells) and then applying energy, e.g. agitating or vibrating the mixture. This results in an array comprising more tightly associated particles, as the agitation is done with sufficient energy to cause weakly-associated beads to fall off (or out, in the case of wells). These sites are then available to bind a different bead. In this way, beads that exhibit a high affinity for the sites are selected. Arrays made in this way have two main advantages as compared to a more static loading: first of all, a higher percentage of the sites can be filled easily, and secondly, the arrays thus loaded show a substantial decrease in bead loss during assays. Thus, in a preferred embodiment, these methods are used to generate arrays that have at least about 50% of the sites filled, with at least about 75% being preferred, and at least about 90% being particularly preferred. Similarly, arrays generated in this manner preferably lose less than about 20% of the beads during an assay, with less than about 10% being preferred and less than about 5% being particularly preferred.

In this embodiment, the substrate comprising the surface with the discrete sites is immersed into a solution comprising the particles (beads, cells, etc.). The surface may comprise wells, as is described herein, or other types of sites on a patterned surface such that there is a differential affinity for the sites. This differential affinity results in a competitive process, such that particles that will associate more tightly are selected. Preferably, the entire surface to be "loaded" with beads is in fluid contact with the solution. This solution is generally a slurry ranging from about 10,000:1 beads:solution (vol:vol) to 1:1. Generally, the solution can comprise any number of reagents, including aqueous buffers, organic solvents, salts, other reagent components, etc. In addition, the solution preferably comprises an excess of beads; that is, there are more beads than sites on the array. Preferred embodiments utilize two-fold to billion-fold excess of beads.

The immersion can mimic the assay conditions; for example, if the array is to be "dipped" from above into a microtiter plate comprising samples, this configuration can be repeated for the loading, thus minimizing the beads that are likely to fall out due to gravity.

Once the surface has been immersed, the substrate, the solution, or both are subjected to a competitive process, whereby the particles with lower affinity can be disassociated from the substrate and replaced by particles exhibiting a higher affinity to the site. This competitive process is done by

the introduction of energy, in the form of heat, sonication, stirring or mixing, vibrating or agitating the solution or substrate, or both.

A preferred embodiment utilizes agitation or vibration. In general, the amount of manipulation of the substrate is minimized to prevent damage to the array; thus, preferred embodiments utilize the agitation of the solution rather than the array, although either will work. As will be appreciated by those in the art, this agitation can take on any number of forms, with a preferred embodiment utilizing microtiter plates comprising bead solutions being agitated using microtiter plate shakers.

The agitation proceeds for a period of time sufficient to load the array to a desired fill. Depending on the size and concentration of the beads and the size of the array, this time may range from about 1 second to days, with from about 1 minute to about 24 hours being preferred.

It should be noted that not all sites of an array may comprise a bead; that is, there may be some sites on the substrate surface which are empty. In addition, there may be some sites that contain more than one bead, although this is not preferred.

In some embodiments, for example when chemical attachment is done, it is possible to attach the beads in a non-random or ordered way. For example, using photoactivatible attachment linkers or photoactivatible adhesives or masks, selected sites on the array may be sequentially rendered suitable for attachment, such that defined populations of beads are laid down.

The arrays of the present invention are constructed such that information about the identity of the candidate agent is built into the array, such that the random deposition of the beads in the fiber wells can be "decoded" to allow identification of the candidate agent at all positions. This may be done in a variety of ways, and either before, during or after the use of the array to detect target molecules.

Thus, after the array is made, it is "decoded" in order to identify the location of one or more of the probes, i.e. each subpopulation of beads, on the substrate surface.

In a preferred embodiment, a selective decoding system is used. In this case, only those microspheres exhibiting a change in the optical signal as a result of the binding of a target analyte are decoded. This is commonly done when the number of "hits", i.e. the number of sites to decode, is generally low. That is, the array is first scanned under experimental conditions in the absence of the target analytes. The sample containing the target analytes is added, and only those locations exhibiting a change in the optical signal are decoded. For example, the beads at either the positive or negative signal locations may be either selectively tagged or released from the array (for example through the use of photocleavable linkers), and subsequently sorted or enriched in a fluorescence-activated cell sorter (FACS). That is, either all the negative beads are released, and then the positive

beads are either released or analyzed in situ, or alternatively all the positives are released and analyzed. Alternatively, the labels may comprise halogenated aromatic compounds, and detection of the label is done using for example gas chromatography, chemical tags, isotopic tags mass spectral tags.

As will be appreciated by those in the art, this may also be done in systems where the array is not decoded; i.e. there need not ever be a correlation of bead composition with location. In this embodiment, the beads are loaded on the array, and the assay is run. The "positives", i.e. those beads displaying a change in the optical signal as is more fully outlined below, are then "marked" to distinguish or separate them from the "negative" beads. This can be done in several ways, preferably using fiber optic arrays. In a preferred embodiment, each bead contains a fluorescent dye. After the assay and the identification of the "positives" or "active beads", light is shown down either only the positive fibers or only the negative fibers, generally in the presence of a light-activated reagent (typically dissolved oxygen). In the former case, all the active beads are photobleached. Thus, upon non-selective release of all the beads with subsequent sorting, for example using a fluorescence activated cell sorter (FACS) machine, the non-fluorescent active beads can be sorted from the fluorescent negative beads. Alternatively, when light is shown down the negative fibers, all the negatives are non-fluorescent and the the postives are fluorescent, and sorting can proceed. The characterization of the attached probe may be done directly, for example using mass spectroscopy.

Alternatively, the identification may occur through the use of identifier moieties ("IMs"), which are similar to IBLs but need not necessarily bind to DBLs. That is, rather than elucidate the structure of the capture probe directly, the composition of the IMs may serve as the identifier. Thus, for example, a specific combination of IMs can serve to code the bead, and be used to identify the agent on the bead upon release from the bead followed by subsequent analysis, for example using a gas chromatograph or mass spectroscope.

Alternatively, rather than having each bead contain a fluorescent dye, each bead comprises a non-fluorescent precursor to a fluorescent dye. For example, using photocleavable protecting groups, such as certain ortho-nitrobenzyl groups, on a fluorescent molecule, photoactivation of the fluorochrome can be done. After the assay, light is shown down again either the "positive" or the "negative" fibers, to distinquish these populations. The illuminated precursors are then chemically converted to a fluorescent dye. All the beads are then released from the array, with sorting, to form populations of fluorescent and non-fluorescent beads (either the positives and the negatives or vice versa).

In an alternate preferred embodiment, the sites of attachment of the beads (for example the wells) include a photopolymerizable reagent, or the photopolymerizable agent is added to the assembled array. After the test assay is run, light is shown down again either the "positive" or the "negative"

fibers, to distinquish these populations. As a result of the irradiation, either all the positives or all the negatives are polymerized and trapped or bound to the sites, while the other population of beads can be released from the array.

In a preferred embodiment, the location of every capture probe is determined using decoder binding ligands (DBLs). As outlined above, DBLs are binding ligands that will either bind to identifier binding ligands, if present, or to the capture probes themselves.

In a preferred embodiment, as outlined above, the DBL binds to the IBL.

In a preferred embodiment, the capture probes are single-stranded nucleic acids and the DBL is a substantially complementary single-stranded nucleic acid that binds (hybridizes) to the capture probe, termed a decoder probe herein. A decoder probe that is substantially complementary to each candidate probe is made and used to decode the array. In this embodiment, the candidate probes and the decoder probes should be of sufficient length (and the decoding step run under suitable conditions) to allow specificity; i.e. each candidate probe binds to its corresponding decoder probe with sufficient specificity to allow the distinction of each candidate probe.

In a preferred embodiment, the DBLs are either directly or indirectly labeled. By "labeled" herein is meant that a compound has at least one element, isotope or chemical compound attached to enable the detection of the compound. In general, labels fall into three classes: a) isotopic labels, which may be radioactive or heavy isotopes; b) magnetic, electrical, thermal; and c) colored or luminescent dyes; although labels include enzymes and particles such as magnetic particles as well. Preferred labels include luminescent labels. In a preferred embodiment, the DBL is directly labeled, that is, the DBL comprises a label. In an alternate embodiment, the DBL is indirectly labeled; that is, a labeling binding ligand (LBL) that will bind to the DBL is used. In this embodiment, the labeling binding ligand-DBL pair can be as described above for IBL-DBL pairs.

Accordingly, the identification of the location of the individual beads (or subpopulations of beads) is done using one or more decoding steps comprising a binding between the labeled DBL and either the IBL or the capture probe (i.e. a hybridization between the capture probe and the decoder probe). After decoding, the DBLs can be removed and the array can be used; however, in some circumstances, for example when the DBL binds to an IBL and not to the capture probe, the removal of the DBL is not required (although it may be desirable in some circumstances). In addition, as outlined herein, decoding may be done either before the array is used to in an assay, during the assay, or after the assay.

In one embodiment, a single decoding step is done. In this embodiment, each DBL is labeled with a unique label, such that the the number of unique tags is equal to or greater than the number of capture

probe (although in some cases, "reuse" of the unique labels can be done, as described herein; similarly, minor variants of candidate probes can share the same decoder, if the variants are encoded in another dimension, i.e. in the bead size or label). For each capture probe or IBL, a DBL is made that will specifically bind to it and contains a unique tag, for example one or more fluorochromes. Thus, the identity of each DBL, both its composition (i.e. its sequence when it is a nucleic acid) and its label, is known. Then, by adding the DBLs to the array containing the bioactive agents under conditions which allow the formation of complexes (termed hybridization complexes when the components are nucleic acids) between the DBLs and either the bioactive agents or the IBLs, the location of each DBL can be elucidated. This allows the identification of the location of each capture probe; the random array has been decoded. The DBLs can then be removed, if necessary, and the target sample applied.

In a preferred embodiment, the number of unique labels is less than the number of unique bioactive agents, and thus a sequential series of decoding steps are used. To facilitate the discussion, this embodiment is explained for nucleic acids, although other types of capture probe and DBLs are useful as well. In this embodiment, decoder probes are divided into **n** sets for decoding. The number of sets corresponds to the number of unique tags. Each decoder probe is labeled in **n** separate reactions with **n** distinct tags. All the decoder probes share the same **n** tags. The decoder probes are pooled so that each pool contains only one of the **n** tag versions of each decoder, and no two decoder probes have the same sequence of tags across all the pools. The number of pools required for this to be true is determined by the number of decoder probes and the **n**. Hybridization of each pool to the array generates a signal at every address. The sequential hybridization of each pool in turn will generate a unique, sequence-specific code for each candidate probe. This identifies the candidate probe at each address in the array. For example, if four tags are used, then 4 X **n** sequential hybridizations can ideally distinguish 4^n sequences, although in some cases more steps may be required. After the hybridization of each pool, the hybrids are denatured and the decoder probes removed, so that the probes are rendered single-stranded for the next hybridization (although it is also possible to hybridize limiting amounts of target so that the available probe is not saturated. Sequential hybridizations can be carried out and analyzed by subtracting pre-existing signal from the previous hybridization).

An example is illustrative. Assuming an array of 16 probe nucleic acids (numbers 1-16), and four unique tags (four different fluors, for example; labels A-D). Decoder probes 1-16 are made that correspond to the probes on the beads. The first step is to label decoder probes 1-4 with tag A, decoder probes 5-8 with tag B, decoder probes 9-12 with tag C, and decoder probes 13-16 with tag D. The probes are mixed and the pool is contacted with the array containing the beads with the attached candidate probes. The location of each tag (and thus each decoder and candidate probe pair) is then determined. The first set of decoder probes are then removed. A second set is added, but this time, decoder probes 1, 5, 9 and 13 are labeled with tag A, decoder probes 2, 6, 10 and 14 are labeled with tag B, decoder probes 3, 7, 11 and 15 are labeled with tag C, and decoder probes 4, 8, 12 and 16 are

labeled with tag D. Thus, those beads that contained tag A in both decoding steps contain candidate probe 1; tag A in the first decoding step and tag B in the second decoding step contain candidate probe 2; tag A in the first decoding step and tag C in the second step contain candidate probe 3; etc. In one embodiment, the decoder probes are labeled in situ; that is, they need not be labeled prior to the decoding reaction. In this embodiment, the incoming decoder probe is shorter than the candidate probe, creating a 5' "overhang" on the decoding probe. The addition of labeled ddNTPs (each labeled with a unique tag) and a polymerase will allow the addition of the tags in a sequence specific manner, thus creating a sequence-specific pattern of signals. Similarly, other modifications can be done, including ligation, etc.

In addition, since the size of the array will be set by the number of unique decoding binding ligands, it is possible to "reuse" a set of unique DBLs to allow for a greater number of test sites. This may be done in several ways; for example, by using some subpopulations that comprise optical signatures. Similarly, the use of a positional coding scheme within an array; different sub-bundles may reuse the set of DBLs. Similarly, one embodiment utilizes bead size as a coding modality, thus allowing the reuse of the set of unique DBLs for each bead size. Alternatively, sequential partial loading of arrays with beads can also allow the reuse of DBLs. Furthermore, "code sharing" can occur as well.

In a preferred embodiment, the DBLs may be reused by having some subpopulations of beads comprise optical signatures. In a preferred embodiment, the optical signature is generally a mixture of reporter dyes, preferably fluoroscent. By varying both the composition of the mixture (i.e. the ratio of one dye to another) and the concentration of the dye (leading to differences in signal intensity), matrices of unique optical signatures may be generated. This may be done by covalently attaching the dyes to the surface of the beads, or alternatively, by entrapping the dye within the bead. The dyes may be chromophores or phosphors but are preferably fluorescent dyes, which due to their strong signals provide a good signal-to-noise ratio for decoding. Suitable dyes for use in the invention include, but are not limited to, fluorescent lanthanide complexes, including those of Europium and Terbium, fluorescein, rhodamine, tetramethylrhodamine, eosin, erythrosin, coumarin, methyl-coumarins, pyrene, Malacite green, stilbene, Lucifer Yellow, Cascade Blue™, Texas Red, and others described in the 6th Edition of the Molecular Probes Handbook by Richard P. Haugland, hereby expressly incorporated by reference.

In a preferred embodiment, the encoding can be accomplished in a ratio of at least two dyes, although more encoding dimensions may be added in the size of the beads, for example. In addition, the labels are distinguishable from one another; thus two different labels may comprise different molecules (i.e. two different fluors) or, alternatively, one label at two different concentrations or intensity.

In a preferred embodiment, the dyes are covalently attached to the surface of the beads. This may be done as is generally outlined for the attachment of the capture probe, using functional groups on the

surface of the beads. As will be appreciated by those in the art, these attachments are done to minimize the effect on the dye.

In a preferred embodiment, the dyes are non-covalently associated with the beads, generally by entrapping the dyes in the pores of the beads.

Additionally, encoding in the ratios of the two or more dyes, rather than single dye concentrations, is preferred since it provides insensitivity to the intensity of light used to interrogate the reporter dye's signature and detector sensitivity.

In a preferred embodiment, a spatial or positional coding system is done. In this embodiment, there are sub-bundles or subarrays (i.e. portions of the total array) that are utilized. By analogy with the telephone system, each subarray is an "area code", that can have the same tags (i.e. telephone numbers) of other subarrays, that are separated by virtue of the location of the subarray. Thus, for example, the same unique tags can be reused from bundle to bundle. Thus, the use of 50 unique tags in combination with 100 different subarrays can form an array of 5000 different bioactive agents. In this embodiment, it becomes important to be able to identify one bundle from another; in general, this is done either manually or through the use of marker beads, i.e. beads containing unique tags for each subarray.

In alternative embodiments, additional encoding parameters can be added, such as microsphere size. For example, the use of different size beads may also allow the reuse of sets of DBLs; that is, it is possible to use microspheres of different sizes to expand the encoding dimensions of the microspheres. Optical fiber arrays can be fabricated containing pixels with different fiber diameters or cross-sections; alternatively, two or more fiber optic bundles, each with different cross-sections of the individual fibers, can be added together to form a larger bundle; or, fiber optic bundles with fiber of the same size cross-sections can be used, but just with different sized beads. With different diameters, the largest wells can be filled with the largest microspheres and then moving onto progressively smaller microspheres in the smaller wells until all size wells are then filled. In this manner, the same dye ratio could be used to encode microspheres of different sizes thereby expanding the number of different oligonucleotide sequences or chemical functionalities present in the array. Although outlined for fiber optic substrates, this as well as the other methods outlined herein can be used with other substrates and with other attachment modalities as well.

In a preferred embodiment, the coding and decoding is accomplished by sequential loading of the microspheres into the array. As outlined above for spatial coding, in this embodiment, the optical signatures can be "reused". In this embodiment, the library of microspheres each comprising a different bioactive agent (or the subpopulations each comprise a different bioactive agent), is divided into a plurality of sublibraries; for example, depending on the size of the desired array and the number

of unique tags, 10 sublibraries each comprising roughly 10% of the total library may be made, with each sublibrary comprising roughly the same unique tags. Then, the first sublibrary is added to the fiber optic bundle comprising the wells, and the location of each bioactive agent is determined, generally through the use of DBLs. The second sublibrary is then added, and the location of each bioactive agent is again determined. The signal in this case will comprise the signal from the "first" DBL and the "second" DBL; by comparing the two matrices the location of each bead in each sublibrary can be determined. Similarly, adding the third, fourth, etc. sublibraries sequentially will allow the array to be filled.

In a preferred embodiment, codes can be "shared" in several ways. In a first embodiment, a single code (i.e. IBL/DBL pair) can be assigned to two or more agents if the target analytes different sufficiently in their binding strengths. For example, two nucleic acid probes used in an mRNA quantitation assay can share the same code if the ranges of their hybridization signal intensities do not overlap. This can occur, for example, when one of the target sequences is always present at a much higher concentration than the other. Alternatively, the two target sequences might always be present at a similar concentration, but differ in hybridization efficiency.

Alternatively, a single code can be assigned to multiple agents if the agents are functionally equivalent. For example, if a set of oligonucleotide probes are designed with the common purpose of detecting the presence of a particular gene, then the probes are functionally equivalent, even though they may differ in sequence. Similarly, if classes of analytes are desired, all probes for different members of a class such as kinases or G-protein coupled receptors could share a code. Similarly, an array of this type could be used to detect homologs of known genes. In this embodiment, each gene is represented by a heterologous set of probes, hybridizing to different regions of the gene (and therefore differing in sequence). The set of probes share a common code. If a homolog is present, it might hybridize to some but not all of the probes. The level of homology might be indicated by the fraction of probes hybridizing, as well as the average hybridization intensity. Similarly, multiple antibodies to the same protein could all share the same code.

In a preferred embodiment, several levels of redundancy are built into the arrays of the invention. Building redundancy into an array gives several significant advantages, including the ability to make quantitative estimates of confidence about the data and signficant increases in sensitivity. Thus, preferred embodiments utilize array redundancy. As will be appreciated by those in the art, there are at least two types of redundancy that can be built into an array: the use of multiple identical sensor elements (termed herein "sensor redundancy"), and the use of multiple sensor elements directed to the same target analyte, but comprising different chemical functionalities (termed herein "target redundancy"). For example, for the detection of nucleic acids, sensor redundancy utilizes of a plurality of sensor elements such as beads comprising identical binding ligands such as probes. Target redundancy utilizes sensor elements with different probes to the same target: one probe may span the

first 25 bases of the target, a second probe may span the second 25 bases of the target, etc. By building in either or both of these types of redundancy into an array, significant benefits are obtained. For example, a variety of statistical mathematical analyses may be done.

In addition, while this is generally described herein for bead arrays, as will be appreciated by those in the art, this techniques can be used for any type of arrays designed to detect target analytes. Furthermore, while these techniques are generally described for nucleic acid systems, these techniques are useful in the detection of other binding ligand/target analyte systems as well.

In a preferred embodiment, sensor redundancy is used. In this embodiment, a plurality of sensor elements, e.g. beads, comprising identical bioactive agents are used. That is, each subpopulation comprises a plurality of beads comprising identical bioactive agents (e.g. binding ligands). By using a number of identical sensor elements for a given array, the optical signal from each sensor element can be combined and any number of statistical analyses run, as outlined below. This can be done for a variety of reasons. For example, in time varying measurements, redundancy can significantly reduce the noise in the system. For non-time based measurements, redundancy can significantly increase the confidence of the data.

In a preferred embodiment, a plurality of identical sensor elements are used. As will be appreciated by those in the art, the number of identical sensor elements will vary with the application and use of the sensor array. In general, anywhere from 2 to thousands may be used, with from 2 to 100 being preferred, 2 to 50 being particularly preferred and from 5 to 20 being especially preferred. In general, preliminary results indicate that roughly 10 beads gives a sufficient advantage, although for some applications, more identical sensor elements can be used.

Once obtained, the optical response signals from a plurality of sensor beads within each bead subpopulation can be manipulated and analyzed in a wide variety of ways, including baseline adjustment, averaging, standard deviation analysis, distribution and cluster analysis, confidence interval analysis, mean testing, etc.

In a preferred embodiment, the first manipulation of the optical response signals is an optional baseline adjustment. In a typical procedure, the standardized optical responses are adjusted to start at a value of 0.0 by subtracting the integer 1.0 from all data points. Doing this allows the baseline-loop data to remain at zero even when summed together and the random response signal noise is canceled out. When the sample is a fluid, the fluid pulse-loop temporal region, however, frequently exhibits a characteristic change in response, either positive, negative or neutral, prior to the sample pulse and often requires a baseline adjustment to overcome noise associated with drift in the first few data points due to charge buildup in the CCD camera. If no drift is present, typically the baseline from the first data point for each bead sensor is subtracted from all the response data for the same bead. If

drift is observed, the average baseline from the first ten data points for each bead sensor is substracted from the all the response data for the same bead. By applying this baseline adjustment, when multiple bead responses are added together they can be amplified while the baseline remains at zero. Since all beads respond at the same time to the sample (e.g. the sample pulse), they all see the pulse at the exact same time and there is no registering or adjusting needed for overlaying their responses. In addition, other types of baseline adjustment may be done, depending on the requirements and output of the system used.

Once the baseline has been adjusted, a number of possible statistical analyses may be run to generate known statistical parameters. Analyses based on redundancy are known and generally described in texts such as Freund and Walpole, Mathematical Statistics, Prentice Hall, Inc. New Jersey, 1980, hereby incorporated by reference in its entirety.

In a preferred embodiment, signal summing is done by simply adding the intensity values of all responses at each time point, generating a new temporal response comprised of the sum of all bead responses. These values can be baseline-adjusted or raw. As for all the analyses described herein, signal summing can be performed in real time or during post-data acquisition data reduction and analysis. In one embodiment, signal summing is performed with a commercial spreadsheet program (Excel, Microsoft, Redmond, WA) after optical response data is collected.

In a preferred embodiment, cummulative response data is generated by simply adding all data points in successive time intervals. This final column, comprised of the sum of all data points at a particular time interval, may then be compared or plotted with the individual bead responses to determine the extent of signal enhancement or improved signal-to-noise ratios.

In a preferred embodiment, the mean of the subpopulation (i.e. the plurality of identical beads) is determined, using the well known Equation 1:

Equation 1

$$\mu = \sum \frac{x_i}{n}$$

In some embodiments, the subpopulation may be redefined to exclude some beads if necessary (for example for obvious outliers, as discussed below).

In a preferred embodiment, the standard deviation of the subpopulation can be determined, generally using Equation 2 (for the entire subpopulation) and Equation 3 (for less than the entire subpopulation):

Equation 2

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{n}}$$

Equation 3

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

As for the mean, the subpopulation may be redefined to exclude some beads if necessary (for example for obvious outliers, as discussed below).

In a preferred embodiment, statistical analyses are done to evaluate whether a particular data point has statistical validty within a subpopulation by using techniques including, but not limited to, t distribution and cluster analysis. This may be done to statistically discard outliers that may otherwise skew the result and increase the signal-to-noise ratio of any particular experiment. This may be done using Equation 4:

Equation 4

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

In a preferred embodiment, the quality of the data is evaluated using confidence intervals, as is known in the art. Confidence intervals can be used to facilitate more comprehensive data processing to measure the statistical validity of a result.

In a preferred embodiment, statistical parameters of a subpopulation of beads are used to do hypothesis testing. One application is tests concerning means, also called mean testing. In this application, statistical evaluation is done to determine whether two subpopulations are different. For example, one sample could be compared with another sample for each subpopulation within an array to determine if the variation is statistically significant.

In addition, mean testing can also be used to differentiate two different assays that share the same code. If the two assays give results that are statistically distinct from each other, then the subpopulations that share a common code can be distinguished from each other on the basis of the assay and the mean test, shown below in Equation 5:

Equation 5

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Furthermore, analyzing the distribution of individual members of a subpopulation of sensor elements may be done. For example, a subpopulation distribution can be evaluated to determine whether the distribution is binomial, Poisson, hypergeometric, etc.

In addition to the sensor redundancy, a preferred embodiment utilizes a plurality of sensor elements that are directed to a single target analyte but yet are not identical. For example, a single target nucleic acid analyte may have two or more sensor elements each comprising a different probe. This adds a level of confidence as non-specific binding interactions can be statistically minimized. When nucleic acid target analytes are to be evaluated, the redundant nucleic acid probes may be overlapping, adjacent, or spatially separated. However, it is preferred that two probes do not compete for a single binding site, so adjacent or separated probes are preferred. Similarly, when proteinaceous target analytes are to be evaluated, preferred embodiments utilize bioactive agent binding agents that bind to different parts of the target. For example, when antibodies (or antibody fragments) are used as bioactive agents for the binding of target proteins, preferred embodiments utilize antibodies to different epitopes.

In this embodiment, a plurality of different sensor elements may be used, with from about 2 to about 20 being preferred, and from about 2 to about 10 being especially preferred, and from 2 to about 5 being particularly preferred, including 2, 3, 4 or 5. Howeve, as above, more may also be used, depending on the application.

As above, any number of statistical analyses may be run on the data from target redundant sensors.

One benefit of the sensor element summing (referred to herein as "bead summing" when beads are used), is the increase in sensitivity that can occur.

Once made, the compositions of the invention find use in a number of applications.

In a preferred embodiment, the probes are used in genetic diagnosis. For example, probes can be made using the techniques disclosed herein to detect target sequences such as the gene for nonpolyposis colon cancer, the BRCA1 breast cancer gene, P53, which is a gene associated with a variety of cancers, the Apo E4 gene that indicates a greater risk of Alzheimer's disease, allowing for

easy presymptomatic screening of patients, mutations in the cystic fibrosis gene, cytochrome p450s or any of the others well known in the art.

In an additional embodiment, viral and bacterial detection is done using the complexes of the invention. In this embodiment, probes are designed to detect target sequences from a variety of bacteria and viruses. For example, current blood-screening techniques rely on the detection of anti-HIV antibodies. The methods disclosed herein allow for direct screening of clinical samples to detect HIV nucleic acid sequences, particularly highly conserved HIV sequences. In addition, this allows direct monitoring of circulating virus within a patient as an improved method of assessing the efficacy of anti-viral therapies. Similarly, viruses associated with leukemia, HTLV-I and HTLV-II, may be detected in this way. Bacterial infections such as tuberculosis, chlamydia and other sexually transmitted diseases, may also be detected.

In a preferred embodiment, the nucleic acids of the invention find use as probes for toxic bacteria in the screening of water and food samples. For example, samples may be treated to lyse the bacteria to release its nucleic acid, and then probes designed to recognize bacterial strains, including, but not limited to, such pathogenic strains as, *Salmonella*, *Campylobacter*, *Vibrio cholerae*, *Leishmania*, enterotoxic strains of *E. coli*, and Legionnaire's disease bacteria. Similarly, bioremediation strategies may be evaluated using the compositions of the invention.

In a further embodiment, the probes are used for forensic "DNA fingerprinting" to match crime-scene DNA against samples taken from victims and suspects.

In an additional embodiment, the probes in an array are used for sequencing by hybridization.

The present invention also finds use as a methodology for the detection of mutations or mismatches in target nucleic acid sequences. For example, recent focus has been on the analysis of the relationship between genetic variation and phenotype by making use of polymorphic DNA markers. Previous work utilized short tandem repeats (STRs) as polymorphic positional markers; however, recent focus is on the use of single nucleotide polymorphisms (SNPs), which occur at an average frequency of more than 1 per kilobase in human genomic DNA. Some SNPs, particularly those in and around coding sequences, are likely to be the direct cause of therapeutically relevant phenotypic variants. There are a number of well known polymorphisms that cause clinically important phenotypes; for example, the apoE2/3/4 variants are associated with different relative risk of Alzheimer's and other diseases (see Cordor et al., Science 261(1993). Multiplex PCR amplification of SNP loci with subsequent hybridization to oligonucleotide arrays has been shown to be an accurate and reliable method of simultaneously genotyping at least hundreds of SNPs; see Wang et al., Science, 280:1077 (1998); see also Schafer et al., Nature Biotechnology 16:33-39 (1998). The compositions of the present invention may easily be substituted for the arrays of the prior art.

In a preferred embodiment, a change in optical signal occurs as a result of the binding of a target analyte that is labeled, either directly or indirectly, with a detectable label, preferably an optical label such as a fluorochrome. Thus, for example, when a nucleic acid target analyte is used, it may be either directly labeled with a fluor, or indirectly, for example through the use of a labeled antibody. Similarly, nucleic acids are easily labeled with fluorochromes, as outlined herein. for example during PCR amplification as is known in the art. Alternatively, upon binding of the target sequences, a hybridization indicator may be used as the label. Hybridization indicators preferentially associate with double stranded nucleic acid, usually reversibly. Hybridization indicators include intercalators and minor and/or major groove binding moieties. In a preferred embodiment, intercalators may be used; since intercalation generally only occurs in the presence of double stranded nucleic acid, only in the presence of target hybridization will the label light up. Thus, upon binding of the target analyte to a bioactive agent, there is a new optical signal generated at that site, which then may be detected.

Furthermore, in some embodiments, a change in the optical signature may be the basis of the optical signal. For example, the interaction of some chemical target analytes with some fluorescent dyes on the beads may alter the optical signature, thus generating a different optical signal.

The assays may be run under a variety of experimental conditions, as will be appreciated by those in the art. A variety of other reagents may be included in the screening assays. These include reagents like salts, neutral proteins, e.g. albumin, detergents, etc which may be used to facilitate optimal protein-protein binding and/or reduce non-specific or background interactions. Also reagents that otherwise improve the efficiency of the assay, such as protease inhibitors, nuclease inhibitors, anti-microbial agents, etc., may be used. The mixture of components may be added in any order that provides for the requisite binding. Various blocking and washing steps may be utilized as is known in the art.

In a preferred embodiment, the methods of the invention are useful in array quality control. Prior to this invention, no methods have been described that provide a positive test of the performance of every probe on every array. Decoding of the array not only provides this test, it also does so by making use of the data generated during the decoding process itself. Therefore, no additional experimental work is required. The invention requires only a set of data analysis algorithms that can be encoded in software.

The quality control procedure can identify a wide variety of systematic and random problems in an array. For example, random specks of dust or other contaminants might cause some sensors to give an incorrect signal-this can be detected during decoding. The omission of one or more agents from multiple arrays can also be detected. An advantage of this quality control procedure is that it can be implemented immediated prior to the assay itself, and is a true functional test of each individual sensor. Therefore any problems that might occur between array assembly and actual use can be

detected. In applications where a very high level of confidence is required, and/or there is a significant chance of sensor failure during the experimental procedure, decoding and quality control can be conducted both before and after the actual sample analysis.

In a preferred embodiment, the arrays can be used to do reagent quality control. In many instances, biological macromolecules are used as reagents and must be quality controlled. For example, large sets of oligonucleotide probes may be provided as reagents. It is typically difficult to perform quality control on large numbers of different biological macromolecules. The approach described here can be used to do this by treating the reagents (formulated as the DBLs) as variable instead of the arrays.

In a preferred embodiment, the methods outlined herein are used in array calibration. For many applications, such as mRNA quantitation, it is desirable to have a signal that is a linear response to the concentration of the target analyte, or, alternatively, if non-linear, to determine a relationship between concentration and signal, so that the concentration of the target analyte can be estimated. Accordingly, the present invention provides methods of creating calibration curves in parallel for multiple beads in an array. The calibration curves can be created under conditions that simulate the complexity of the sample to be analyzed. Each curve can be constructed independently of the others (e.g. for a different range of concentrations), but at the same time as all the other curves for the array. Thus, in this embodiment, the sequential decoding scheme is implemented with different concentrations being used as the code "labels", rather than different fluorophores. In this way, signal as a response to concentration can be measured for each bead. This calibration can be carried out just prior to array use, so that every probe on every array is individually calibrated as needed.

In a preferred embodiment, the methods of the invention can be used in assay development as well. Thus, for example, the methods allow the identification of good and bad probes; as is understood by those in the art, some probes do not function well because they do not hybridize well, or because they cross-hybridize with more than one sequence. These problems are easily detected during decoding. The ability to rapidly assess probe performance has the potential to greatly reduce the time and expense of assay development.

Similarly, in a preferred embodiment, the methods of the invention are useful in quantitation in assay development. Major challenges of many assays is the ability to detect differences in analyte concentrations between samples, the ability to quantitate these differences, and to measure absolute concentrations of analytes, all in the presence of a complex mixture of related analytes. An example of this problem is the quantitation of a specific mRNA in the presence of total cellular mRNA. One approach that has been developed as a basis of mRNA quantitation makes use of a multiple match and mismatch probe pairs (Lockhart et al., 1996), hereby incorporated by reference in its entirety. While this approach is simple, it requires relatively large numbers of probes. In this approach, a quantitative response to concentration is obtained by averaging the signals from a set of different

probes to the gene or sequence of interest. This is necessary because only some probes respond quantitatively, and it is not possible to predict these probes with certainty. In the absence of prior knowledge, only the average response of an appropriately chosen collection of probes is quantitative. However, in the present invention, this can be applied generally to nucleic acid based assays as well as other assays. In essence, the approach is to identify the probes that respond quantitatively in a particular assay, rather than average them with other probes. This is done using the array calibration scheme outlined above, in which concentration-based codes are used. Advantages of this approach include: fewer probes are needed; the accuracy of the measurement is less dependent on the number of probes used; and that the response of the sensors is known with a high level of certainty, since each and every sequence can be tested in an efficient manner. It is important to note that probes that perfom well are chosen empirically, which avoids the difficulties and uncertainties of predicting probe performance, particularly in complex sequence mixtures. In contrast, in experiments described to date with ordered arrays, relatively small numbers of sequences are checked by performing quantitative spiking experiments, in which a known mRNA is added to a mixture.

All references cited herein are incorporated by reference in their entirety.